

TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-883-49

Query Match 2.3%; Score 18; DB 1; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 8
US-08-832-877-49/c
Sequence 49, Application US/08832877
Patent No. 5840506
GENERAL INFORMATION:

APPLICANT: Giordano, Antonio
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 116
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: US/08/832,877

CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US2
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-877-49

Query Match 2.3%; Score 18; DB 2; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 9
US-08-832-877-49

Sequence 16, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 2.3%; Score 18; DB 3; Length 152331;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 8501 AAGGCAAAAAA 8484

RESULT 10
US-08-858-767-21
Sequence 21, Application US/08858767
Patent No. 5837468
GENERAL INFORMATION:
APPLICANT: WANG, Xun
APPLICANT: DUVICK, Jonathan P.
APPLICANT: BRIGGS, Steven P.
TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
NUMBER OF SEQUENCES: 39
TITLE OF INVENTION: METHOD
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,767
FILING DATE: 19-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/481,687
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/325/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

us-09-817-318-1.oli.rni

```

;
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 28 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
US-08-858-767-21

```

Qy 764 aggcataaaataaaaaa 780
|||
Db 7 AGGCATAAATAAAAAA 23

Patent No. 5837468
GENERAL INFORMATION:
APPLICANT: WANG, Xun
APPLICANT: DUVICK, Jonathan P.
APPLICANT: BRIGGS, Steven P.
TITLE OF INVENTION: PCR-BASED
TITLE OF INVENTION: METHOD

```

1 ZIP: 20007-5109
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/858,710
16
17 FILING DATE: 19-MAY-1997
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1
```

```

; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-858-767-22

```

Qy 764 aggcaccccccccaaa 780
db 7 AGGCACCCCCCAAAA 23

US-08-858-767-23
 : Sequence 23, Application US/08858767
 : Patent No. 5937468
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: WANG, Xun
 : APPLICANT: DUVICK, Jonathan P.
 : APPLICANT: BRIGGS, Steven P.
 : TITLE OF INVENTION: PCR-BASED CD
 : TITLE OF INVENTION: METHOD
 : NUMBER OF SEQUENCES: 39
 : CORRESPONDENCE ADDRESS:
 :

```

, COUNTRY: USA
, ZIP: 20007-5109
,
, COMPUTER READABLE FORM:
,
, MEDIUM TYPE: Floppy disk
, COMPUTER: IBM PC compatible
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: PatentIn Release #1.0
,

```

APPLICATION NUMBER: US 08/481,687
FILING DATE: 07-JUN-1995

```

; INFORMATION FOR SEQ ID NO: 23:
;-----
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 28 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
US-08-858-767-23

```

Qy 764 aggcataaaataaaaaa 780
 |||||
 Db 7 AGGCATAAATAAAAAA 23

```

? GENERAL INFORMATION:
? APPLICANT: WANG, Xun
? APPLICANT: DUVICK, Jonathan P.
? APPLICANT: BRIGGS, Steven P.
? TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
? TITLE OF INVENTION: METHOD
? NUMBER OF SEQUENCES: 39
? CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 22:53:53 ; Search time 64.69 seconds
(without alignments)
2961.727 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgttcttctaag.....taaggcaaaaaaaaaaaaaa 780

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	2.6	588	US-09-385-982-275	Sequence 275, App
C 2	20	2.6	176373	US-09-128-155-17	Sequence 17, Appl
C 3	19	2.4	6143	US-08-612-521-3	Sequence 3, Appl
C 4	19	2.4	7122	US-09-318-448-2	Sequence 2, Appl
C 5	18	2.3	776	US-09-372-422A-43	Sequence 43, Appl
C 6	18	2.3	3171	PCT-US95-09261-1	Sequence 1, Appl
C 7	18	2.3	4576	US-08-832-883-49	Sequence 49, Appl
C 8	18	2.3	4576	US-08-832-877-49	Sequence 49, Appl
C 9	18	2.3	152331	US-09-128-155-16	Sequence 16, Appl
C 10	17	2.2	28	US-08-858-767-21	Sequence 21, Appl
C 11	17	2.2	28	US-08-858-767-22	Sequence 22, Appl
C 12	17	2.2	28	US-08-858-767-23	Sequence 23, Appl
C 13	17	2.2	28	US-08-863-028-21	Sequence 21, Appl
C 14	17	2.2	28	US-08-863-028-22	Sequence 22, Appl
C 15	17	2.2	28	US-08-863-028-23	Sequence 23, Appl
C 16	17	2.2	418	US-08-642-274D-46	Sequence 46, Appl
C 17	17	2.2	418	US-08-952-014C-46	Sequence 46, Appl
C 18	17	2.2	545	US-09-328-111-506	Sequence 506, App
C 19	17	2.2	668	US-08-516-545-1	Sequence 516, Appl
C 20	17	2.2	854	US-08-684-672-25	Sequence 25, Appl
C 21	17	2.2	854	US-08-684-672-26	Sequence 26, Appl
C 22	17	2.2	1134	US-09-248-335-29	Sequence 29, Appl
C 23	17	2.2	1235	US-08-672-850-5	Sequence 5, Appl
C 24	17	2.2	1630	PCT-US92-01196-6	Sequence 6, Appl
C 25	17	2.2	1640	US-08-781-250-1	Sequence 1, Appl
C 26	17	2.2	1697	US-09-346-408-5	Sequence 5, Appl
C 27	17	2.2	1853	US-09-439-313-369	Sequence 369, App

ALIGNMENTS

RESULT 1

US-09-385-982-275/c

; Sequence 275, Application US/09385982

; Patent No. 6262334

; GENERAL INFORMATION:

; APPLICANT: ENDEGE, WILSON O., ET AL.

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCDNA-260XX

; CURRENT APPLICATION NUMBER: US/09/385,982

; CURRENT FILING DATE: 1999-08-30

; EARLIER APPLICATION NUMBER: 09/328,111

; EARLIER FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: 60/117,393

; EARLIER FILING DATE: 1999-01-27

; EARLIER APPLICATION NUMBER: 60/098,639

; EARLIER FILING DATE: 1998-08-31

; NUMBER OF SEQ ID NOS: 544

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 275

; LENGTH: 588

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(588)

; OTHER INFORMATION: n = A,T,C or G

US-09-385-982-275

Query Match 2.6%; Score 20; DB 4; Length 588;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 taaaggcaaaaaaaaaaaaaa 780

|||||

Db 30 TAAAGGCAAAAAAAAAAAAAA 11

RESULT 2

US-09-128-155-17/c

; Sequence 17, Application US/09128155

; Patent No. 6117654

; GENERAL INFORMATION:

; APPLICANT: Pan, Yang

; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY

; FILE REFERENCE: 09404/052001

; CURRENT APPLICATION NUMBER: US/09/128,155

; CURRENT FILING DATE: 1998-08-03

EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(176373)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match 2.6%; Score 20; DB 3; Length 176373;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780

Db 93715 TAAAGCAAAAAA 93696

RESULT 3

US-08-612-521-3/c
Sequence 3, Application US/08612521
Patent No. 5786463

GENERAL INFORMATION:

APPLICANT: Peery, Robert B
APPLICANT: Skatrud, Paul L
APPLICANT: Thorne, Susan J
TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:

ADDRESS: Patent Division/AEH
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/612,521

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hamilton, Amy E
REGISTRATION NUMBER: 33,894
REFERENCE/DOCKET NUMBER: X-9693
TELEPHONE: 317-276-3169
TELEFAX: 317-276-1294

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 6143 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

FEATURE:

NAME/KEY: exon

LOCATION: 541..1072

FEATURE:

NAME/KEY: exon

LOCATION: 1130..1395

FEATURE:

NAME/KEY: exon
LOCATION: 1443..1872
FEATURE:
NAME/KEY: exon
LOCATION: 1925..2428
FEATURE:
NAME/KEY: exon
LOCATION: 2479..2863
FEATURE:
NAME/KEY: exon
LOCATION: 2913..3009
FEATURE:
NAME/KEY: exon
LOCATION: 3076..3379
FEATURE:
NAME/KEY: exon
LOCATION: 3438..3561
FEATURE:
NAME/KEY: exon
LOCATION: 3614..3776
FEATURE:
NAME/KEY: exon
LOCATION: 3831..3966
FEATURE:
NAME/KEY: exon
LOCATION: 4016..4244
FEATURE:
NAME/KEY: exon
LOCATION: 4291..4438
FEATURE:
NAME/KEY: exon
LOCATION: 4487..4743
FEATURE:
NAME/KEY: exon
LOCATION: 4793..4958
FEATURE:
NAME/KEY: exon
LOCATION: 5006..5113
FEATURE:
NAME/KEY: exon
LOCATION: 5162..5500
FEATURE:
NAME/KEY: exon
LOCATION: 5553..5588
US-08-612-521-3

Query Match 2.4%; Score 19; DB 1; Length 6143;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 aaaggcaaaaaa 780

Db 3037 AAAGCAAAAAA 3019

RESULT 4

US-09-318-448-2/c
Sequence 2, Application US/09318448
Patent No. 6210950

GENERAL INFORMATION:

APPLICANT: Johnson, William G.
APPLICANT: Steenroos, Edward S.
TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
FILE REFERENCE: 601-1-057
CURRENT APPLICATION NUMBER: US/09/318,448
CURRENT FILING DATE: 1999-05-25
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2

LENGTH: 7122

TYPE: DNA

; ORGANISM: Homo sapiens
US-09-318-448-2

Query Match 2.4%; Score 19; DB 4; Length 7122;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 aaagcagcaaaaaaa 780
|||||

Db 4104 AAAGGCAAAAAA 4086

RESULT 5

US-09-372-422A-43

; Sequence 43, Application US/09372422A

; Patent No. 6313375

; GENERAL INFORMATION:

; APPLICANT: Rudolf Jung

; APPLICANT: Francois Barriau

; TITLE OF INVENTION: Maize Aquaporins and Uses Thereof

; FILE REFERENCE: 0919

; CURRENT APPLICATION NUMBER: US/09/372,422A

; CURRENT FILING DATE: 1999-08-11

; PRIOR APPLICATION NUMBER: US 60/098,692

; PRIOR FILING DATE: 1998-08-31

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 43

; LENGTH: 776

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (20)...(341)

US-09-372-422A-43

Query Match

Best Local Similarity 100.0%; Pred. No. 40; Length 776;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 aaggcaaaaaaa 780
|||||

Db 747 aaggcaaaaaaa 764

RESULT 6

PCT-US95-09261-1

; Sequence 1, Application PC/TUS9509261

; GENERAL INFORMATION:

; APPLICANT:

; APPLICANT: NAME: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM

; APPLICANT: STREET: 201 West 7th Street

; APPLICANT: CITY: Austin

; APPLICANT: STATE: Texas

; APPLICANT: COUNTRY: United States of America

; APPLICANT: POSTAL CODE: 78701

; APPLICANT: TELEPHONE NO: (512)499-4462

; APPLICANT: TELEFAX: (512)499-4523

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE EXPRESSION OF

; TITLE OF INVENTION: A BONE AND PROSTATE DERIVED GROWTH FACTOR

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/09261

; FILING DATE: CONCURRENTLY HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/283,701

; FILING DATE: 01-AUG-1994

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: SERTICH, GARY J.

; REGISTRATION NUMBER: 34,430

; REFERENCE/DOCKET NUMBER: UTFC422P--

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (713) 789-2679

; TELEX: 79-0924

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3171 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "DNA"

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 694..2310

PCT-US95-09261-1

Query Match

Best Local Similarity 100.0%; Pred. No. 36; Score 18; DB 5; Length 3171;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 763 aaggcaaaaaaa 780
|||||

Db 3153 AAGGCAAAAAA 3170

RESULT 7

US-08-832-883-49/c

; Sequence 49, Application US/08832883

; Patent No. 5807681

; GENERAL INFORMATION:

; APPLICANT: Giordano, Antonio

; APPLICANT: Baldi, Alphonso

; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS

; TITLE OF INVENTION: OF CANCER

; NUMBER OF SEQUENCES: 115

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.

; STREET: Suite 1800 Two Penn Center Plaza

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/832,883

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Monaco, Daniel A.

; REGISTRATION NUMBER: 30,480

; REFERENCE/DOCKET NUMBER: 8321-13 US1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 568-8383

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 21:37:15 ; Search time 2133.92 Seconds
(without alignments)
7649.161 Million cell updates/sec

Title: US-09-817-318-1
Perfect score: 780
Sequence: 1 aattactgttcttcttaag.....taaggcaaaaaaaaaa 780

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	780	100.0	780	6	AX260297	Sequence
2	352	45.1	167021	9	AC079414	Homo sapi
3	34	4.4	34	6	AX260319	Sequence
4	23	2.9	23	6	AX260317	Sequence
5	23	2.9	1470	9	BC017168	Homo sapi
6	23	2.9	73590	2	AC022759	Homo sapi
7	23	2.9	180336	2	AC021731	Homo sapi
8	23	2.9	189036	9	AC006043	Homo sapi
9	22	2.8	22	6	AX260318	Sequence
10	22	2.8	37714	3	CET11F9	274042 Caenorhabdi
11	22	2.8	95008	2	AC087654	Homo sapi
12	22	2.8	110000	2	AC013622_0	AC013622 Mus muscu
13	22	2.8	110000	2	AC013623_0	AC013623 Mus muscu
14	22	2.8	113729	9	AF260011	Homo sapi
15	22	2.8	115865	9	AL358944	Human DNA
16	22	2.8	129108	9	AC011482	Homo sapi
17	22	2.8	136862	9	AC005007	Homo sapi
18	22	2.8	160566	2	AC079247	Homo sapi
19	22	2.8	181532	2	AC020700	Homo sapi
20	22	2.8	181996	2	AC068749	Homo sapi
21	22	2.8	184213	9	AC073539	Homo sapi
22	22	2.8	225009	2	AC093366	Mus muscu
23	22	2.8	252278	2	AC090126	Mus muscu
24	22	2.8	340000	9	HS21C048	AL163248 Homo sapi
25	21	2.7	733	9	HS3332860	AJ332860 Homo sapi
26	21	2.7	1234	9	AF264785	Homo sapi
27	21	2.7	1282	8	AF209988	Batrachos
28	21	2.7	1471	9	AK000415	Homo sapi
29	21	2.7	49183	9	HS0139A3	AL109916 Human DNA
30	21	2.7	55227	2	AC100331	Mus muscu
31	21	2.7	61020	6	AX251546	AX251546 Sequence
32	21	2.7	62868	2	AC016799	Homo sapi
33	21	2.7	71203	9	AL160251	Human DNA
34	21	2.7	90656	2	AC095790	Rattus no
35	21	2.7	91388	9	AC005994	Homo sapi
36	21	2.7	144223	9	AC062015	Homo sapi
37	21	2.7	152327	2	AC026295	Homo sapi
38	21	2.7	156443	2	AC091910	Homo sapi
39	21	2.7	157257	2	AC073580	Homo sapi
40	21	2.7	162990	2	AC024707	Homo sapi
41	21	2.7	163039	2	AC010997	Homo sapi
42	21	2.7	164297	2	AC012138	Homo sapi
43	21	2.7	165412	2	AC010811	Homo sapi
44	21	2.7	169169	2	AC092527	Papio cyn
45	21	2.7	169612	2	AC027393	Homo sapi

ALIGNMENTS

RESULT	1	AX260297	Sequence 1 from Patent WO0172780.	780 bp	DNA	linear	PAT 26-OCT-2001
LOCUS	AX260297	Sequence 1 from Patent WO0172780.					
DEFINITION	AX260297	Sequence 1 from Patent WO0172780.					
ACCESSION	AX260297	Sequence 1 from Patent WO0172780.					
VERSION	AX260297.1	GI:16509264					
KEYWORDS	human.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
TITLE	1 (sites)						
JOURNAL	Salceda, S., Hu, P., Recipon, H. and Cafferkey, R.						
FEATURES	Compositions and methods of diagnosing, monitoring, staging, imaging and treating mammary gland cancer						
Source	Patent: WO 0172780-A 1 04-OCT-2001;						
	diadexus, Inc. (US)						
	Location/Qualifiers						
	1..780						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
BASE COUNT	267 a	115 c	119 g	279 t			
ORIGIN							

Query Match 100.0%; Score 780; DB 6; Length 780;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 aattactgttctcttaagtaagccttacaccctactaaaatgtgatcaaaatttat 60
Db 1 AATTACTGTCTCTTAAAGTAAGGCCCTACACCCTACTAAAATGTGATCAAAATTTAT 60

Qy 61 tatgaatagatgaaagcgtagctataaaattatgagagtaagtttatatttatc 120
Db 61 TATGATAGATGAAGCGTGTAGCTATAAAATATGACAGTAAGTTTATTTATATTATC 120

Qy 121 caaatgtagtccataatagcataatagcaacttcactaaatcttagataaaaaaagaat 180
Db 121 CAATGTAGTTTCATAATAGCATATAGCAACTTCACCTAAATCTTAGAATAAAAAATGAAT 180

Qy 181 aaaatgtaattttttggaggaatggttaattttttctacaaaattgtgtgacagcttt 240
Db 181 AAAATGTTAATTTTGGAGGAATGGTTAATTTTCTACAAAATGTGTGACAGCTTT 240

Qy 241 acagaccttactcttcacaaatgacttgaacattacacacacacagaggctcgtgttac 300
Db 241 ACAGACCTTACTCTTCAATGACTTGAACATTAACATCACAAGAGGCTCGTGTTC 300

Qy 301 aaaagaatagtcagaacttcagaaatttttgacagtgactcttttcaaccctttaac 360
Db 301 AAAAGAATAGTCAGAACTTCATGAATTTTGTGACAGTGACTCTTTCTAACCTTTAATC 360

Qy 361 caaatatttaagtgtccatgctctctctttatcacaactcaatttttaactagttttct 420
Db 361 CAATATATTTAAGTGCCATGCTCTCTCTTTATCCAACTCAATTTGTTAACTAGTTTCT 420

Qy 421 tctgtgagttcccttgctataaattgaagcagttctctgaaatcaccccaactgatttta 480
Db 421 TCTGTGAGTTCCCTTGCTATAATTGAAGCAGTTCTCTGAATCACCACCAACTGATTTA 480

Qy 481 tgaagcccatgcttttgaagagatttgacactcggttgcaatctattacattgact 540
Db 481 TGAAGCCCATGCTTTTGAAGAGATTGTGCACTTCGGCTTTGCAATCTATTATTCATTGACT 540

Qy 541 gtactgtcattgtctagatgttgactatcagttaggaacatacaaaaagatataga 600
Db 541 GTACTTGCACTGTATGTGATGTGACTATCAGTTAGGACAATCAAAAAGATATTAGA 600

Qy 601 taatggcgaggataaatacagaagtactgtcaataaacaagaatttatgtttatgggtatt 660
Db 601 TAATGGCGAGGATAAATCAGAAGTTACTGTCAATAACAAGATTATGTTTATGGGTATT 660

Qy 661 ttatagtgataaattcattactagcaattcatatcatgttttaattctctctggttgt 720
Db 661 TTATAGTGATAAATTCATTACTAGCAATTCATATCATGTGTTTAAATCTCTCTGGTTGT 720

Qy 721 aatatgtgactctggagactcaaatataaattatgtgtaagggcaaaaaaataaaa 780
Db 721 AATATGGTGACTCTGAGACTCAAAATATTAAATATTGTGTGAAGGCCAAAAAATAAAA 780

```

RESULT 2

AC079414/c 167021 bp DNA linear PRI 04-JAN-2002
 LOCUS Homo sapiens chromosome 16 clone RP11-358L22, complete sequence.
 DEFINITION Homo sapiens
 ACCESSION AC079414
 VERSION AC079414.5 GI:18057080
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 (bases 1 to 167021)
 DOE Joint Genome Institute.
 Sequencing of Human Chromosome 16

REFERENCE
 AUTHORS
 TITLE

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 167021)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 167021)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Jan 4, 2002 this sequence version replaced gi:17976466.
 Sequence Quality Assessment:

This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

FEATURES
 Source

Location/Qualifiers
 1..167021
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-358L22"

BASE COUNT 49239 a 36098 c 35465 g 46219 t
 ORIGIN

Query Match 45.1%; Score 352; DB 9; Length 167021;
 Best Local Similarity 99.4%; Pred. No. 3.3e-166;
 Matches 622; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

Qy 140 cataatgacactcactaaatcttagaataaaaaatgaataaaatgttaatttttga 199
Db 94271 CATAATAGCAACTTCACATAATCTTAGAATAAAAAATCAATAAATGTTAATTTTGA 94212

Qy 200 ggaatggttaattttttacaaaattgtgacagctttacagaccttactcttca 259
Db 94211 GGAATGGTTAATTTTCTACAAAATGTGTGACAGCTTTACAGACTTACTCTTCACA 94152

Qy 260 attgacttgaacattacacacacagagggtcctgtttacaaaagaataagtcagaact 319
Db 94151 ATTGACTTGAACATTAACATCACAAGAGGGTCTCTGTGTGACAAAAGTAAGTCAAGACT 94092

Qy 320 tcatgaattttgacagtgactcttttcaacccttaaccataatatttaagtctc 379
Db 94091 TCATGAATTTTGTGACAGTGACTCTTTCTAACCTTTTAATCCAAATATATTTAAGTGTC 94032

Qy 380 atogtctctttatccaactcatttgttaactagtttctctctgtgagttcttgcct 439
Db 94031 ATCGTCGTCCTTTATCAACTCATTTGTTAACTAGTTTCTTCTGTGAGTTCCTTTGCCT 93972

Qy 440 ataattgaagcagttctctgaaatcaccccaactgattttatgaaagccatgctttgg 499
Db 93971 ATAATTGAAGCAGTTCTCTGAATCACCACCAACTGATTTTATGAAAGCCCATGCTTTGG 93912

Qy 500 aaagatttgcaactcggttgcattcttattacattgactgactgattgattgt 559
Db 93911 AAAGATTGCACTTCGGCTTTGCAATCTATTATACATTGACTTGTGATTGTCT 93852

```

```
QY 560 agatgttgactatcagttaggaacatcaaaagatatagataatggcagggataaac 619
|||||
Db 93851 AGATGTTGACTATCAGTTAGGACANTCAAAAATATATT-CATAATGGCAGGATAAATC 93793

QY 620 agaagttactgtcaatacaaaagtattgttttattgggtattttatagtgataaaattcat 679
|||||
Db 93792 AGAAGTTACTGTCATAAACAAAGTTATGTTTATGGCTATTTTATAGTGATAAAATTCAT 93733

QY 680 tactgacgaatttcataatcattttaaattctctctgttgaataatgtgactctggaga 739
|||||
Db 93732 TACTGACGAATTCATATCAATGTTTAAATCTCTCTGTTGTTAAATGTGACTCTGGAGA 93673

QY 740 tcaaaatattaaatattggtgtaaaag 765
|||||
Db 93672 CTCAAAATATTAAATATTGGGTAAAG 93647

RESULT 3
AX260319/c
LOCUS AX260319 34 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 23 from Patent WO0172780.
ACCESSION AX260319
VERSION AX260319.1 GI:16509286
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (sites)
AUTHORS Salceda,S., Hu,P., Recipon,H. and Cafferkey,R.
TITLE Compositions and methods of diagnosing, monitoring, staging,
JOURNAL imaging and treating mammary gland cancer
diabex, Inc. (US)
Patent: WO 0172780-A 23 04-OCT-2001;
diabex, Inc. (US)
FEATURES
source
1. .34
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic"
BASE COUNT 17 a 6 c 6 g 5 t
ORIGIN
|||||
QY 414 gttttctctgtgagttccttgcataattga 447
|||||
Db 34 GTTTCTCTGTGAGTTCCTTTGCTATAATTGA 1

RESULT 4
AX260317
LOCUS AX260317 23 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 21 from Patent WO0172780.
ACCESSION AX260317
VERSION AX260317.1 GI:16509284
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (sites)
AUTHORS Salceda,S., Hu,P., Recipon,H. and Cafferkey,R.
TITLE Compositions and methods of diagnosing, monitoring, staging,
JOURNAL imaging and treating mammary gland cancer
diabex, Inc. (US)
Patent: WO 0172780-A 21 04-OCT-2001;
diabex, Inc. (US)
FEATURES
source
1. .23
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic"
```

```
BASE COUNT 4 a 9 c 1 g 9 t
ORIGIN
|||||
Query Match 2.9%; Score 23; DB 6; Length 23;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 ccacgtctctcttaccacaaat 400
|||||
Db 1 CCATCGTCTCTCTTATCCAACT 23

RESULT 5
BC017168
LOCUS BC017168 1470 bp mRNA linear PRI 09-NOV-2001
DEFINITION Homo sapiens, clone MGC:9517 IMAGE:3907072, mRNA, complete cds.
ACCESSION BC017168
VERSION BC017168.1 GI:16877877
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1470)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 15 Row: b Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.
FEATURES
Location/Qualifiers
1. .1470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:9517 IMAGE:3907072"
/tissue_type="Uterus, leiomyosarcoma"
/clone_lib="NIH-MGC_71"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
195. .1118
/codon_start=1
/product="Unknown (protein for MGC:9517)"
/protein_id="AAH17168.1"
/db_xref="GI:16877878"
/translation="MTYKAGSGGNLRKLDGNELDLSLSDINPEVPKELAAALPKATIL
DLSCKNLTTLPSCDFGLTHLVKLDLSKNKLOQLPADFGRLVNLQHLDLNNKVLTPV
SFAQLKNLWLDLNDPLVLAKEKCKOCOKANKVQHMKAQVADQDREKQ
RRLEVEREAEKKREAKQRAKEAOFERELRKRKAEKEKERRRRKEYDALKAAKREQKKPK"
CDS
```

KEANOAPKSGSRKPPKPRKHTKRSWVLKLLLLLLLFGVAGLVACRVTELOQQOPL
 CTSVNTIYDNAOGLRRRIEILQWVLOIIDSQ"

BASE COUNT
 ORIGIN

364 a 377 c 460 g 269 t

Query Match

Best Local Similarity 2.9%; Score 23; DB 9; Length 1470;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 758 gtgttaagcgcaaaaaa 780

Db 1422 GTGTAAGCGCAAAAAA 1444

RESULT 6

AC022759 73590 bp DNA linear HTG 13-JUL-2000
 Homo sapiens clone RP11-273K8, LOW-PASS SEQUENCE SAMPLING.

AC022759
 AC022759.2 GI:9153403

HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 73590)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-273K8

Unpublished

2 (bases 1 to 73590)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavskiy,L., Boukhgaltier,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeArelano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPeeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,F.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.

Direct Submission

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6922562.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5832

Center clone name: 273_K_8

* NOTE: This record contains 84 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that

* the record is updated, the accession number will
 * be preserved.

1 759: contig of 759 bp in length
 760 859: gap of 100 bp
 860 1639: contig of 780 bp in length
 1640 1739: gap of 100 bp
 1740 2522: contig of 783 bp in length
 2523 2622: gap of 100 bp
 2623 3400: contig of 778 bp in length
 3401 3500: gap of 100 bp
 3501 4277: contig of 777 bp in length
 4278 4377: gap of 100 bp
 4378 5145: contig of 768 bp in length
 5146 5245: gap of 100 bp
 5246 6016: contig of 771 bp in length
 6017 6116: gap of 100 bp
 6117 6892: contig of 776 bp in length
 6893 6992: gap of 100 bp
 6993 7779: contig of 787 bp in length
 7780 7879: gap of 100 bp
 7880 8652: contig of 773 bp in length
 8653 8752: gap of 100 bp
 8753 9544: contig of 792 bp in length
 9545 9644: gap of 100 bp
 9645 10410: contig of 766 bp in length
 10411 10510: gap of 100 bp
 10511 11294: contig of 784 bp in length
 11295 11394: gap of 100 bp
 11395 12150: contig of 756 bp in length
 12151 12250: gap of 100 bp
 12251 13045: contig of 795 bp in length
 13046 13145: gap of 100 bp
 13146 13940: contig of 795 bp in length
 13941 14040: gap of 100 bp
 14041 14807: contig of 767 bp in length
 14808 14907: gap of 100 bp
 14908 15661: contig of 754 bp in length
 15662 15761: gap of 100 bp
 15762 16537: contig of 776 bp in length
 16538 16637: gap of 100 bp
 16638 17415: contig of 778 bp in length
 17416 17515: gap of 100 bp
 17516 18314: contig of 799 bp in length
 18315 18414: gap of 100 bp
 18415 19202: contig of 788 bp in length
 19203 19302: gap of 100 bp
 19303 20078: contig of 776 bp in length
 20079 20178: gap of 100 bp
 20179 20967: contig of 789 bp in length
 20968 21067: gap of 100 bp
 21068 21860: contig of 793 bp in length
 21861 21960: gap of 100 bp
 21961 22746: contig of 786 bp in length
 22747 22846: gap of 100 bp
 22847 23633: contig of 787 bp in length
 23634 23733: gap of 100 bp
 23734 24513: contig of 780 bp in length
 24514 24613: gap of 100 bp
 24614 25387: contig of 774 bp in length
 25388 25487: gap of 100 bp
 25488 26252: contig of 765 bp in length
 26253 26352: gap of 100 bp
 26353 27115: contig of 763 bp in length
 27116 27215: gap of 100 bp
 27216 27968: contig of 753 bp in length
 27969 28068: gap of 100 bp
 28069 28835: contig of 767 bp in length
 28836 28935: gap of 100 bp
 28936 29723: contig of 788 bp in length
 29724 29823: gap of 100 bp
 29824 30581: contig of 758 bp in length
 30582 30681: gap of 100 bp
 30682 31460: contig of 779 bp in length

31461 31560: gap of 100 bp
31561 32345: contig of 785 bp in length
32346 32445: gap of 100 bp
32446 32440: contig of 795 bp in length
32441 33340: gap of 100 bp
33341 34124: contig of 784 bp in length
34125 34224: gap of 100 bp
34225 35003: contig of 779 bp in length
35004 35103: gap of 100 bp
35104 35873: contig of 770 bp in length
35874 35973: gap of 100 bp
35974 36746: contig of 773 bp in length
36747 36846: gap of 100 bp
36847 37613: contig of 767 bp in length
37614 37713: gap of 100 bp
37714 38487: contig of 774 bp in length
38488 38587: gap of 100 bp
38588 39362: contig of 775 bp in length
39363 39462: gap of 100 bp
39463 40254: contig of 792 bp in length
40255 40354: gap of 100 bp
40355 41141: contig of 787 bp in length
41142 41241: gap of 100 bp
41242 42033: contig of 792 bp in length
42034 42133: gap of 100 bp
42134 42915: contig of 782 bp in length
42916 43015: gap of 100 bp
43016 43794: contig of 779 bp in length
43795 43894: gap of 100 bp
43895 44667: contig of 773 bp in length
44668 44767: gap of 100 bp
44768 45556: contig of 789 bp in length
45557 45656: gap of 100 bp
45657 46433: contig of 777 bp in length
46434 46533: gap of 100 bp
46534 47292: contig of 759 bp in length
47293 47392: gap of 100 bp
47393 48175: contig of 783 bp in length
48176 48275: gap of 100 bp
48276 49063: contig of 788 bp in length
49064 49163: gap of 100 bp
49164 49981: contig of 818 bp in length
49982 50081: gap of 100 bp
50082 50849: contig of 768 bp in length
50850 50949: gap of 100 bp
50950 51740: contig of 791 bp in length
51741 51840: gap of 100 bp
51841 52638: contig of 798 bp in length
52639 52738: gap of 100 bp
52739 53519: contig of 781 bp in length
53520 53619: gap of 100 bp
53620 54360: contig of 741 bp in length
54361 54460: gap of 100 bp
54461 55236: contig of 776 bp in length
55237 55336: gap of 100 bp
55337 56088: contig of 752 bp in length
56089 56188: gap of 100 bp
56189 56978: contig of 790 bp in length
56979 57078: gap of 100 bp
57079 57862: contig of 784 bp in length
57863 57962: gap of 100 bp
57963 58757: contig of 795 bp in length
58758 58857: gap of 100 bp
58858 59640: contig of 783 bp in length
59641 59740: gap of 100 bp
59741 60504: contig of 764 bp in length
60505 60604: gap of 100 bp
60605 61370: contig of 766 bp in length
61371 61470: gap of 100 bp
61471 62237: contig of 767 bp in length
62238 62337: gap of 100 bp

Query Match

2.9%; Score 23; DB 2; Length 73590;

Best Local Similarity 100.0%; Pred. NO. 0.82; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0;
Qy 173 aaatgaataaaatgtaattttt 195
|||||
Db 30380 AAATGAATAAAATCTTAATTTT 30402
RESULT 7
AC021731/c 180336 bp DNA linear HTG 16-MAR-2000
LOCUS Homo sapiens chromosome 11 clone RP11-693K15 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 13 unordered pieces.
AC021731
AC021731.3 GI:7249233
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 180336)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-693K15
Unpublished
2 (bases 1 to 180336)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6982105.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5996
Center clone name: 693_K.15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 171270 bases at least Q40
Consensus quality: 176121 bases at least Q30
Consensus quality: 177872 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 179136; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 5507: contig of 5507 bp in length
 5508 5607: gap of 100 bp
 5608 12210: contig of 6603 bp in length
 12211 12310: gap of 100 bp
 12311 19212: contig of 6902 bp in length
 19213 19312: gap of 100 bp
 19313 28832: contig of 7520 bp in length
 28833 26932: gap of 100 bp
 26933 37970: contig of 11038 bp in length
 37971 38070: gap of 100 bp
 38071 47850: contig of 9780 bp in length
 47851 47950: gap of 100 bp
 47951 60814: contig of 12864 bp in length
 60815 60914: gap of 100 bp
 60915 75929: contig of 15015 bp in length
 75930 76029: gap of 100 bp
 76030 87989: contig of 11960 bp in length
 87990 88089: gap of 100 bp
 88090 104142: contig of 16053 bp in length
 104143 104242: gap of 100 bp
 104243 124453: contig of 20211 bp in length
 124454 124553: gap of 100 bp
 124554 146046: contig of 21493 bp in length
 146047 146146: gap of 100 bp
 146147 180336: contig of 34190 bp in length.

FEATURES

source

1. 180336
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-693K15"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 5507
 /note="assembly_fragment"
 5608. 12210
 /note="assembly_fragment"
 12311. 19212
 /note="assembly_fragment"
 19313. 26832
 /note="assembly_fragment"
 26933. 37970
 /note="assembly_fragment"
 38071. 47850
 /note="assembly_fragment"
 47951. 60814
 /note="assembly_fragment"
 60915. 75929
 /note="assembly_fragment"
 76030. 87989
 /note="assembly_fragment"
 88090. 104142
 /note="assembly_fragment"
 104243. 124453
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 124554. 146046
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right
 146147. 180336
 /note="assembly_fragment"

BASE COUNT 57400 a 31804 c 32636 g 57294 t 1202 others
 ORIGIN

Query:Match 2.98; Score 23; DB 2; Length 180336;
 Best Local Similarity 100.0%; Pred. No. 0.7;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 758 gtgtaaaggcaaaaaa 780
 Db 29872 GTGTAAAGGCCAAAAA 29850
 RESULT 8
 LOCUS AC006043 189036 bp DNA linear PRI 21-DEC-1999
 DEFINITION Homo sapiens BAC clone RP11-538D15 from 7q11.23-q21.1, complete
 sequence.
 ACCESSION AC006043
 VERSION AC006043.1 GI:4156139
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 189036)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 REFERENCE 2 (bases 1 to 189036)
 AUTHORS Kozlowicz, A. and Eldred, J.
 TITLE The sequence of Homo sapiens BAC clone RP11-538D15
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 189036)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (22-NOV-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 189036)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 189036)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (20-FEB-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 6 (bases 1 to 189036)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jan 14, 1999 this sequence version replaced gi:3907457.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics

 Center project name: H_NH0538D15

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.


```

RESULT 9
AX260318/c
LOCUS AX260318 22 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 22 from Patent WO0172780.
ACCESSION AX260318
VERSION AX260318.1 GI:15509285
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
          artificial sequence.
REFERENCE 1 (sites)
AUTHORS Salceda,S., Hu,P., Recipon,H. and Cafferkey,R.
TITLE Compositions and Methods of diagnosing, monitoring, staging,
        imaging and treating mammary gland cancer
JOURNAL Patent: WO 0172780-A 22 04-OCT-2001;
        diaDexus, Inc. (US)
FEATURES
    source
        1..22
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="Synthetic"
BASE COUNT 5 a 3 c 7 g 7 t
ORIGIN
Query Match 2.8%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 448 agcagttctctgaaatcaccca 469
      |||||
Db 22 AGCAGTCTCTGAAATCACC CA 1

RESULT 10
CET11F9/c
LOCUS CET11F9 37714 bp DNA linear INV 24-JAN-2002
DEFINITION Caenorhabditis elegans cosmid T11F9, complete sequence.
ACCESSION Z74042
VERSION Z74042.1 GI:1403259
KEYWORDS HTG; 3-oxoacyl [acyl-carrier protein] reductase; Amino acid
          permease; Collagen; G-protein coupled receptor; Human mRNA
          (K1AA0062) like; Zinc metalloprotease.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
          Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (sites)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
        investigating biology. The C. elegans Sequencing Consortium
        Science 282 (5396), 2012-2018 (1998)
JOURNAL 99069613
MEDLINE
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 37714)
AUTHORS Lennard,N.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-1996) Nematode Sequencing Project, Sanger
        Institute, Hinxton, Cambridge CB10 1SA, England and Department of
        Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
        Jcs@sanger.ac.uk or rw@nematoe.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using
        predictions from Genefinder (P. Green, U. Washington), and other
        available information.
        Current sequence finishing criteria for the C. elegans genome
        sequencing consortium are that all bases are either sequenced
        unambiguously on both strands, or on a single strand with both a
        dye primer and dye terminator reaction, from distinct subclones.
        Exceptions are indicated by an explicit note.
        This sequence is the entire insert of clone T11F9. The true right
        end of clone F5787 is at 15157 in this sequence. The start of this
        sequence (1..105) overlaps with the end of sequence Z74037.
        The end of this sequence (37615..37714) overlaps with the start of

```

sequence AL031623.
For a graphical representation of this sequence and its analysis
see:- <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=T11F9>
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.

FEATURES

source

```

1..37714
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="v"
/clone="T11F9"
join(1579..1633,1708..1834,2118..2325,2374..2514,
2562..3071,3251..3332,3416..3531,3746..3841)
/genes="T11F9.1"
join(1579..1633,1708..1834,2118..2325,2374..2514,
2562..3071,3251..3332,3416..3531,3746..3841)
/genes="T11F9.1"
/note="similar to G-protein coupled receptor"
/codon_start=1
/protein_id="CAA98536.1"
/db_xref="GI:3879696"
/db_xref="SPTREMBL:Q22405"
/translation="MDYESTSFQDQSTATNDQECAYSEPVYVKKERFWVAIFGTAVC
LINLENTSEFLMKKSRSSHMLYLALLAFEDVVMALAYIPLMSLNLVDYDKSV
VLLRAWFAYLMPITVSHIAMTASSFLVAAASLERIVITCHPTKNRSLRNRWIAAF
AIFLGTCTCKESQLEYEMIEYLPQCMTMREYQNLNLSALARYEYGVWRVHTVITL
IPFSLAFINIRIVLSKNEFKLHSTKLSDAKRSASVRNATRMVIVFYLLSNV
LNVIIWEYIDMDLTOQFETFYMEAVDVSLTIFGALRLPIYVICOANLRKEFF
AOLTMLSLNSYTGFLSSSTKNNTIEDMDIDRTTIDISGDGVLFVIRPSDNPLA"
TVIRKNSLRHMGFGFSMEKRFKQLNKRKILKLIENQFIRPSDNPLA"
join(6815..6821,6883..7171,7384..7675,7847..8512,
8586..8732)
/genes="T11F9.2"
join(6815..6821,6883..7171,7384..7675,7847..8512,
8586..8732)
/genes="T11F9.2"
/note="Similarity to Human mRNA for novel protein
(K1AA0062) (TR:G505102)"
/codon_start=1
/protein_id="CAA98527.2"
/db_xref="GI:15718225"
/db_xref="SPTREMBL:Q22395"
/translation="MTSHGIGFAIVSGCSFAPLIGLLPCLSKSLYPRIMTFLVAVG
IGALSGSTIFMLPQAFHLSFEHFHTKSLIILCALYAFETVDMLOYLLEFRRR
QTKRIHASTIASLMTPTTKRRDNGHNTTEETIVPSEPTITRQLVDPDLNHHFRFR
ESELSDVETEQEKEMAEALNDLEMLTNNVLARTFTRRVAVSGDLDDIEFRSP
KLSSHNTNGNTSOFLDVINNEFHRRMTPLSSRPGSPVTINVEEPKESYEMKQSEKPG
LNHQDSDMSVSIRVVEKKEVIEPAAMEVASVAYMIFGSAANFNVDGMSKGAFSDN
LLRGLSIGIAVISQOFPOELGTALILVKSLGLKLTLLFNVPVIVLSPLGFSIGWMLD
SVDDSYDEYIFAISGGMYMYIFLGLTLPETRESTNELIKENLAESILVSIQACGLF
GTFWYFMSRVNSVDF"
complement(join(8969..9135,9187..9535,9579..9675,
9947..10533,11041..11163))
/genes="T11F9.4"
complement(join(8969..9135,9187..9535,9579..9675,
9947..10533,11041..11163))
/genes="T11F9.4"
/note="Similarity to Schistosoma amino acid permease
(TR:G407047), contains similarity to Pfam domain: PF00324
(Amino acid permease), Score=-206.3, E-value=1.7e-05, N=1"
/codon_start=1
/protein_id="CAA98529.1"
/db_xref="GI:3879689"
/db_xref="SPTREMBL:Q22397"
/translation="MPDSDRSQKMLGALSYIVGNIVSGSIFITPTSIENVNSTOI
QGQVEYLDFTSEWAKKLGFLSLIILMFNFFSLKTFVQFSLASLAKIAATLLII
ITGFYLLIFKHWKQNEEPKSGSNMFPFNALFAGLFYDGDWDLNFCAGAEAFVVP
RTVPLSLIIGNTGCTIVYAVNAVYSIVLSPTEIASNAVAIDFANKTLCAAAAFVVP
MVAIIIGLSTMTFMSASRYLOAVSROGHIPSAISGAPNSDPRVALVHILAIATV
SFLGDPKLINYVAFQWSQRAFTMSALLYLRINGRPHRPHRIQLPIMPILFLVCT

```

gene

CDS

gene

CDS

gene

CDS


```

SVVSIIDDFKSSAVGLGILLGLIIFIWVMDRALPSSHTFRNATHVINNEESTKFM
QIIFVPRVGDDEMKNAIGAESESEKVPAYKISPTGNGOFKCTRM"
complement(join(12965..13249,13297..13639,13692..14035))
/gene="t1lf9.3"
CDS
complement(join(12965..13249,13297..13639,13692..14035))
/gene="t1lf9.3"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=123.8,
E-value=1e-33, N=1"
/codon_start=1
/protein_id="CAA98528.1"
/db_xref="GI:3879688"
/db_xref="SPTREMBL:Q22396"
/translation="MFRDAINLENHTCLKFEYNEAETAVIRKNGKCGSVLYGMHAG
EVQDLTLVNCASFQTAVEIMHAGIAHQARSRDDYLIVDSTNSNDGIENENLV
PDYGSVMLYARDPSDKRIPIDPEYFTMGSLRVAFYDMVLLNKFYGCNDRPKKL
DCKNGYQNPANCEECLECDGFNGQLCDQHEGVYLEAKKWDASGRNNYRKGIEIN
TMPEYTFALTAEGETIEVRIITKLSGFCFCQCDYNGVELKYKTDRIIVSLVCCDN
DNLWNRSTNNPFTIAKYGNNRTPHFEEYRIPGNATAAPEENN"
join(14572..14782,14825..15212,15259..15457)
/gene="t1lf9.5"
CDS
join(14572..14782,14825..15212,15259..15457)
/gene="t1lf9.5"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=92.5,
E-value=2.8e-24, N=1"
/codon_start=1
/protein_id="CAA98533.1"
/db_xref="GI:3879693"
/db_xref="SPTREMBL:Q22401"
/translation="MNYFIFFFMHIAVLNFIYFRFSNGNKIVMRVGGSPETKRLEKSK
ROALRMDNPPRPGTINVFYDEORFEDNSRATVLRAMEKISNHTCIKFSPPDARIKL
RIVSDKGOCAIGRVGDOOYLSEFTSCYSVGSASELHVIGFLSHORADREYLLK
NLQRLNDWFOHYYKKYLDQWLVIPDYGISIMQYHSDNDEYKPKNSYFRTWGSQI
PSFYDLMTNETYQSCGEQEIQNCKNRTPNPGNSECNPLGFG"
join(16710..16914,16958..17355)
/gene="t1lf9.6"
CDS
join(16710..16914,16958..17355)
/gene="t1lf9.6"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=1.1,
E-value=1.2e-09, N=1"
/codon_start=1
/protein_id="CAA98530.1"
/db_xref="GI:3879690"
/db_xref="SPTREMBL:Q22398"
/translation="MKSFFILSIQECYKGDIVARIGRNVAEKIGGARHRROVLIR
GDEERRHKWFNTVHYFYEFNFDTVKESILRAMELISNHTCIKFSTPESEKSIKM
ESDSTTIACVAGVQVRENQLESFNSDCYSAGVAVHELHISLGFIAHQRSDRDQYLE
FKKNDLENOTYEQYKIWEYOELVYPYDVGSMQYPNEE"
join(17911..17989,18042..18276,18320..18390,18435..18507,
18564..18677,18728..18905)
/gene="t1lf9.7"
CDS
join(17911..17989,18042..18276,18320..18390,18435..18507,
18564..18677,18728..18905)
/gene="t1lf9.7"
/notes="contains similarity to Pfam domain: PF01400
(Astacin (Peptidase family M12A)), Score=1.1,
E-value=1.2e-09, N=1"
/codon_start=1
/protein_id="CAA98531.1"
/db_xref="GI:3879691"
/db_xref="SPTREMBL:Q22399"
/translation="MEDESSAKYPYRESMPRAERYSMKALQDEWRHLFTTAEGENNI
HNPKIPCRDNDPDMPCNESESEITGDREVFVSHNEEDEDQESSQICESEP
OELPKLCRTDKHFKELPSGLNITKDLSDLSGLYLPOTLLTGRIDLNDREFAVEV
EAEKSKELLRRRRRPLARQFSHALSRAMANLRYHYVDVTEKIERCPMRNSCKGH
LKWEDMRKTRAOCKNFKDPSUATVEDMMK"
join(19342..19659,19708..19857)
/gene="t1lf9.8"
CDS
join(19342..19659,19708..19857)
/gene="t1lf9.8"
/notes="similar to zinc metalloprotease"
/codon_start=1
/protein_id="CAA98532.1"
/db_xref="GI:3879692"

```

```

/db_xref="SPTREMBL:Q22400"
/translation="MANTMGSAIVAFYDYLMIINKYYECSCANNLCKNKHGYPNPSNCS
QCNCYPGCGADCDSORAEPCATEQWVTLISLDAGYRYLENKOKLQVDFIYQF
LWIMAFANKTQIRVRFVGEKCLPGCIRGVEIKTWNEDPLTSPRUCCEETS"
20147..20290
/gene="t1lf9.15"
/notes="U4-1-RNA
similar to U4-1 snRNA"
20147..20290
/gene="t1lf9.15"
complement(22696..22833)
/gene="t1lf9.14"
complement(22696..22833)
/gene="t1lf9.14"
/codon_start=1
/protein_id="CAA98534.1"
/db_xref="GI:3879694"

Query Match      2.8%  Score 22;  DB 3;  Length 37714;
Best Local Similarity 100.0%;  Pred. No. 2.9;
Matches 22;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 166 gaataaaaaatgaataaatgt 187
|||||
Db 34725 GAATAAAAAATGAATAAAATGCT 34704

RESULT 11
AC087654
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP4-562D20 map 11, WORKING DRAFT .
SEQUENCE, 6 unordereded pieces.
AC087654
AC087654.2 GI:14971248
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 95008)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP4-562D20
Unpublished
REFERENCE
2 (bases 1 to 95008)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B., Brown,A.,
Camaraata,J., Campopiano,A., Choepel,I., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Headford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karakas,A., Larocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,K., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 20, 2001 this sequence version replaced gi:12229429.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L12119
 Center clone name: 562_D_20
 ----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 92202 bases at least Q40
 Consensus quality: 93559 bases at least Q30
 Consensus quality: 94148 bases at least Q20
 Insert size: 115000; agarose-fp
 Quality coverage: 5.8 in Q20 bases; agarose-fp
 Quality coverage: 7.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1416: contig of 1416 bp in length
 1417 1516: gap of 100 bp
 1517 33891: contig of 32375 bp in length
 33892 33991: gap of 100 bp
 33992 41314: contig of 7323 bp in length
 41315 41414: gap of 100 bp
 41415 66165: contig of 24751 bp in length
 66166 66265: gap of 100 bp
 66266 92375: contig of 26110 bp in length
 92376 92475: gap of 100 bp
 92476 95008: contig of 2533 bp in length.

FEATURES

source
 1. 95008
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP4-562D20"
 /clone_lib="RPC1 Human PAC library 4"
 1. 1416
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 1517. 33891
 /note="assembly_fragment"
 33992. 41314
 /note="assembly_fragment"
 41415. 66165
 /note="assembly_fragment"
 66266. 92375
 /note="assembly_fragment"
 92476. 95008
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 BASE COUNT 30236 a 18609 c 17991 g 27672 t 500 others
 ORIGIN

Query Match 2.8%; Score 22; DB 2; Length 95008;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 329 tttagacagtactctttctaa 350
 |||||||
 Db 60637 TTTGACAGTACTCTTTCTAA 60658

RESULT 12
 AC013622_0
 WPCOMMENT

Sequence split into 5 fragments LOCUS AC013622 Accession AC013622
 Fragment Name Begin End
 AC013622_0 1 110000
 AC013622_1 100001 210000
 AC013622_2 200001 310000
 AC013622_3 300001 410000
 AC013622_4 400001 456313

LOCUS AC013622 Mus musculus clone RP23-232H18, *** SEQUENCING IN PROGRESS ***
 DEFINITION Mus musculus clone RP23-232H18, DNA linear HTG 30-MAY-2000
 unsorted pieces.

ACCESSION AC013622
 VERSION AC013622.4 GI:8102534
 KEYWORDS HTG; PHASE1.
 SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 436313)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Mus musculus, clone RP23-232H18
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 436313)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Headford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 30, 2000 this sequence version replaced gi:7329381.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L3193
 Center clone name: 232_H_18

TITLE
 JOURNAL

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 52 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2045: contig of 2045 bp in length
 2046 2145: gap of 100 bp
 2146 4521: contig of 2376 bp in length
 4522 4621: gap of 100 bp
 4622 8024: contig of 3403 bp in length
 8025 8124: gap of 100 bp
 8125 11099: contig of 2975 bp in length
 11100 11199: gap of 100 bp

* 11200 13868: contig of 2669 bp in length
* 13869 13968: gap of 100 bp
* 13969 18430: contig of 4462 bp in length
* 18431 18530: gap of 100 bp
* 18531 21696: contig of 3166 bp in length
* 21697 21796: gap of 100 bp
* 21797 25741: contig of 3945 bp in length
* 25742 25841: gap of 100 bp
* 25842 29357: contig of 3516 bp in length
* 29358 29457: gap of 100 bp
* 29458 32719: contig of 3262 bp in length
* 32720 32819: gap of 100 bp
* 32820 35995: contig of 3176 bp in length
* 35996 36095: gap of 100 bp
* 36096 41396: contig of 5301 bp in length
* 41397 41496: gap of 100 bp
* 41497 45153: contig of 3657 bp in length
* 45154 45253: gap of 100 bp
* 45254 49318: contig of 4065 bp in length
* 49319 49418: gap of 100 bp
* 49419 54714: contig of 5296 bp in length
* 54715 54814: gap of 100 bp
* 54815 58285: contig of 3471 bp in length
* 58286 58385: gap of 100 bp
* 58386 63447: contig of 5062 bp in length
* 63448 63547: gap of 100 bp
* 63548 67966: contig of 4419 bp in length
* 67967 68066: gap of 100 bp
* 68067 73138: contig of 5072 bp in length
* 73139 73238: gap of 100 bp
* 73239 77301: contig of 4063 bp in length
* 77302 77401: gap of 100 bp
* 77402 82081: contig of 4680 bp in length
* 82082 82181: gap of 100 bp
* 82182 86719: contig of 4538 bp in length
* 86720 86819: gap of 100 bp
* 86820 92113: contig of 5294 bp in length
* 92114 92213: gap of 100 bp
* 92214 96761: contig of 4548 bp in length
* 96762 96861: gap of 100 bp
* 96862 101458: contig of 4597 bp in length
* 101459 101558: gap of 100 bp
* 101559 107167: contig of 5609 bp in length
* 107168 107267: gap of 100 bp
* 107268 113324: contig of 6057 bp in length
* 113325 113424: gap of 100 bp
* 113425 119681: contig of 6257 bp in length
* 119682 119781: gap of 100 bp
* 119782 126056: contig of 6275 bp in length
* 126057 126156: gap of 100 bp
* 126157 132230: contig of 6074 bp in length
* 132231 132330: gap of 100 bp
* 132331 140500: contig of 8170 bp in length
* 140501 140600: gap of 100 bp
* 140601 149110: contig of 8510 bp in length
* 149111 149210: gap of 100 bp
* 149211 156954: contig of 7744 bp in length
* 156955 157054: gap of 100 bp
* 157055 163961: contig of 6907 bp in length
* 163962 164061: gap of 100 bp
* 164062 172377: contig of 8316 bp in length
* 172378 172477: gap of 100 bp
* 172478 181342: contig of 8865 bp in length
* 181343 181442: gap of 100 bp
* 181443 189265: contig of 7823 bp in length
* 189266 189365: gap of 100 bp
* 189366 197225: contig of 7860 bp in length
* 197226 197325: gap of 100 bp
* 197326 206982: contig of 9657 bp in length
* 206983 207082: gap of 100 bp
* 207083 215973: contig of 8891 bp in length
* 215974 216073: gap of 100 bp
* 216074 227424: contig of 11351 bp in length

* 227425 227524: gap of 100 bp
* 227525 237208: contig of 9684 bp in length
* 237209 237308: gap of 100 bp
* 237309 247157: contig of 9849 bp in length
* 247158 247257: gap of 100 bp
* 247258 261074: contig of 13817 bp in length
* 261075 261174: gap of 100 bp
* 261175 274135: contig of 12961 bp in length
* 274136 274235: gap of 100 bp
* 274236 289816: contig of 15581 bp in length
* 289817 289916: gap of 100 bp
* 289917 305346: contig of 15430 bp in length
* 305347 305446: gap of 100 bp
* 305447 320002: contig of 14556 bp in length
* 320003 320102: gap of 100 bp
* 320103 339946: contig of 19844 bp in length
* 339947 340046: gap of 100 bp
* 340047 366637: contig of 26591 bp in length
* 366638 366737: gap of 100 bp
* 366738 396109: contig of 29372 bp in length
* 396110 396209: gap of 100 bp
* 396210 436313: contig of 40104 bp in length.

FEATURES

source

1. .436313
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="Rp23-232H18"

/clone_lib="RP23-232H18"

BASE COUNT 122368 a 96120 c 95329 g 116149 t 6347 others
ORIGIN

Query Match

2.8%; Score 22; DB 2; Length 110000;

Best Local Similarity 100.0%; Pred. No. 2.4; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780

|||||

Db 97865 TGTAAAGGCAAAAAA 97886

RESULT 13

AC013623_0/c

WPCOMMENT

Sequence split into 5 fragments LOCUS AC013623 Accession AC013623

Fragment Name	Begin	End
AC013623_0	1	110000
AC013623_1	100001	210000
AC013623_2	200001	310000
AC013623_3	300001	410000
AC013623_4	400001	502225

LOCUS AC013623 Mus musculus clone RP23-6518, *** SEQUENCING IN PROGRESS ***, 54
DEFINITION

unordered pieces.

AC013623

AC013623.5 GI:8099784

KEYWORDS HTG: HTGS_PHASE1.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 502225)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Mus musculus, clone RP23-6518

Unpublished

2 (bases 1 to 502225)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,

Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczyk, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McSwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (13-NOV-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 28, 2000 this sequence version replaced gi:7329382.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L3195
Center clone name: 65_I_8

* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 15255: contig of 15255 bp in length
* 15256 15355: gap of 100 bp
* 15356 41823: contig of 26468 bp in length
* 41824 41923: gap of 100 bp
* 41924 69700: contig of 27777 bp in length
* 69701 69800: gap of 100 bp
* 69801 88128: contig of 18328 bp in length
* 88129 88228: gap of 100 bp
* 88229 120914: contig of 32686 bp in length
* 120915 121014: gap of 100 bp
* 121015 146910: contig of 25896 bp in length
* 146911 147010: gap of 100 bp
* 147011 186035: contig of 39025 bp in length
* 186036 186135: gap of 100 bp
* 186136 220111: contig of 33976 bp in length
* 220112 220211: gap of 100 bp
* 220212 222624: contig of 2413 bp in length
* 222625 222724: gap of 100 bp
* 222725 224842: contig of 2118 bp in length
* 224843 224942: gap of 100 bp
* 224943 227053: contig of 2111 bp in length
* 227054 227153: gap of 100 bp
* 227154 229756: contig of 2603 bp in length
* 229757 229856: gap of 100 bp
* 229857 232049: contig of 2193 bp in length
* 232050 232149: gap of 100 bp
* 232150 235053: contig of 2904 bp in length
* 235054 235153: gap of 100 bp
* 235154 237754: contig of 2601 bp in length
* 237755 237854: gap of 100 bp
* 237855 240561: contig of 2707 bp in length
* 240562 240661: gap of 100 bp
* 240662 244335: contig of 3674 bp in length
* 244336 244435: gap of 100 bp
* 244436 246925: contig of 2490 bp in length
* 246926 247025: gap of 100 bp
* 247026 249480: contig of 2455 bp in length
* 249481 249580: gap of 100 bp
* 249581 252655: contig of 3075 bp in length
* 252656 252755: gap of 100 bp
* 252756 256196: contig of 3441 bp in length
* 256197 256296: gap of 100 bp

* 256297 259587: contig of 3291 bp in length
* 259588 259687: gap of 100 bp
* 259688 261942: contig of 2255 bp in length
* 261943 262042: gap of 100 bp
* 262043 265148: contig of 3106 bp in length
* 265149 265248: gap of 100 bp
* 265249 268367: contig of 3119 bp in length
* 268368 268467: gap of 100 bp
* 268468 272612: contig of 4145 bp in length
* 272613 272712: gap of 100 bp
* 272713 276750: contig of 4038 bp in length
* 276751 276850: gap of 100 bp
* 276851 280557: contig of 3707 bp in length
* 280558 280657: gap of 100 bp
* 280658 284201: contig of 3544 bp in length
* 284202 284301: gap of 100 bp
* 284302 288323: contig of 4022 bp in length
* 288324 288423: gap of 100 bp
* 288424 293056: contig of 4633 bp in length
* 293057 293156: gap of 100 bp
* 293157 298341: contig of 5185 bp in length
* 298342 298441: gap of 100 bp
* 298442 303389: contig of 4948 bp in length
* 303390 303489: gap of 100 bp
* 303490 309662: contig of 6173 bp in length
* 309663 309762: gap of 100 bp
* 309763 313152: contig of 3390 bp in length
* 313153 313252: gap of 100 bp
* 313253 318191: contig of 4939 bp in length
* 318192 318291: gap of 100 bp
* 318292 322669: contig of 4378 bp in length
* 322670 322769: gap of 100 bp
* 322770 329137: contig of 6368 bp in length
* 329138 329237: gap of 100 bp
* 329238 336394: contig of 7157 bp in length
* 336395 336494: gap of 100 bp
* 336495 341574: contig of 5080 bp in length
* 341575 341674: gap of 100 bp
* 341675 347177: contig of 5503 bp in length
* 347178 347277: gap of 100 bp
* 347278 353035: contig of 5758 bp in length
* 353036 353135: gap of 100 bp
* 353136 361314: contig of 8179 bp in length
* 361315 361414: gap of 100 bp
* 361415 370891: contig of 9477 bp in length
* 370892 370991: gap of 100 bp
* 370992 379658: contig of 8667 bp in length
* 379659 379758: gap of 100 bp
* 379759 387395: contig of 7637 bp in length
* 387396 387495: gap of 100 bp
* 387496 396045: contig of 8550 bp in length
* 396046 396145: gap of 100 bp
* 396146 410181: contig of 14036 bp in length
* 410182 410281: gap of 100 bp
* 410282 426416: contig of 16135 bp in length
* 426417 426516: gap of 100 bp
* 426517 441179: contig of 14663 bp in length
* 441180 441279: gap of 100 bp
* 441280 455205: contig of 13926 bp in length
* 455206 455305: gap of 100 bp
* 455306 468105: contig of 12800 bp in length
* 468106 468205: gap of 100 bp
* 468206 481736: contig of 13531 bp in length
* 481737 481836: gap of 100 bp
* 481837 502225: contig of 20389 bp in length.

FEATURES
Location/Qualifiers
1. 502225

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-6518"

BASE COUNT 134208 a 114407 c 114480 g 132045 t 7085 others
ORIGIN

Query Match 2.8%; Score 22; DB 2; Length 110000;
 Best Local Similarity 100.0%; Pred No. 2.4;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780
 |||||
 Db 34977 TGTAAGGCAAAAAA 34956

RESULT 14
 AF260011/c
 LOCUS
 DEFINITION Homo sapiens chromosome 21 clone PAC H08866 map 21q22.1, complete sequence.
 ACCESSION AF260011
 VERSION AF260011.2 GI:13621230
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 113729)
 Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhaber,M.B., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuzumi,N., Nordstieck,G., Antonarakis,S.E., Minoshima,S., Shimizu,N., Desario,A., Hornisch,K., Brandt,P., Schafke,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Riesselmann,L., Dagand,E., Haaf,T., Wehrmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.Laure.
 The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium
 Nature 405 (6784), 311-319 (2000)
 20289799
 2 (bases 1 to 113729)
 Taudien,S., Dagand,E., Hildmann,T., Delabar,J., Blechschmidt,K., Schillhaber,M., Baumgart,C., Dette,M., Menzel,U., Yaspo,M.-L. and Rosenthal,A.
 Direct Submission
 Submitted (20-APR-2000) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
 3 (bases 1 to 113729)
 Taudien,S., Dagand,E., Hildmann,T., Delabar,J., Blechschmidt,K., Schillhaber,M., Baumgart,C., Dette,M., Menzel,U., Yaspo,M.-L. and Rosenthal,A.
 Direct Submission
 Submitted (13-APR-2001) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 COMMENT On Apr 13, 2001 this sequence version replaced gi:7677558.
 FEATURES
 Location/Qualifiers
 1. 113729
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /clone="PAC H08866"
 37225 a 21999 c 20722 g 33783 t
 BASE COUNT
 ORIGIN

Query Match 2.8%; Score 22; DB 9; Length 113729;
 Best Local Similarity 100.0%; Pred No. 2.4;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 gttattttatattatccaa 124
 |||||

Db 95438 GTTTATTTATTTATCCAA 95417

RESULT 15
 AL358944/c
 LOCUS
 DEFINITION Human DNA sequence from clone RP4-562D20 on chromosome 11 contains the FSHB gene for follicle stimulating hormone beta polypeptide, ESTs, STSS and GSSs, complete sequence.
 ACCESSION AL358944
 VERSION AL358944.12 GI:14727299
 KEYWORDS HTG; follicle stimulating hormone; FSHB.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 115865)
 Martin,S.
 Direct Submission
 Submitted (21-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On May 31, 2001 this sequence version replaced gi:13446452.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; SW; SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP4-562D20 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP4-562D20 The true left end of clone RP3-414F14 is at 22396 in this sequence. The true right end of clone RP4-815A16 is at 1704 in this sequence. This sequence was generated by the Sanger Centre from a human chromosome 11 bacterial clone contig constructed by Gavin B. Niederfuhr A, Schumacher N, Hummerich H, Little PF, Gessler M. Genome Res 1999 Nov;9(11):1074-86.
 Nov;9(11):1074-86.
 FEATURES
 Location/Qualifiers
 1. 115865
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP4-562D20"
 /clone_lbb="RPCI-4"
 45. .630
 /note="MER76 repeat: matches 32. .654 of consensus"
 635. .931
 /note="LIM4 repeat: matches 4749. .5048 of consensus"
 935. .1895
 /note="LIM4 repeat: matches 2953. .3955 of consensus"
 1911. .2120
 /note="AluJb repeat: matches 80. .289 of consensus"
 2123. .2582
 /note="LIM4 repeat: matches 2507. .2959 of consensus"
 2600. .2766
 /note="LIM4 repeat: matches 2212. .2372 of consensus"
 4209. .4570
 /note="THE1B repeat: matches 1. .364 of consensus"

```

repeat_region 5323..5607
/feature="AluY repeat: matches 1. .284 of consensus"
repeat_region 5629..5707
/feature="L2 repeat: matches 2627. .2702 of consensus"
repeat_region 6036..6548
/feature="MLR1B repeat: matches 2. .505 of consensus"
misc_feature 6295..6512
/feature="Single clone region. Assembly confirmed by restriction digest data."
repeat_region 6716..7101
/feature="MLR1A2 repeat: matches 1. .374 of consensus"
repeat_region 9191..9745
/feature="MLR2A repeat: matches 1. .452 of consensus"
misc_feature 9906..10692
/feature="Sequence confirmed by AC068749 sequenced by W1BR."
repeat_region 11176..11247
/feature="36 copies 2 mer tc 84% conserved"
repeat_region 11692..11777
/feature="MLR1J repeat: matches 116. .200 of consensus"
misc_feature 11714..11788
/feature="Sequence confirmed by AC068749 sequenced by W1BR."
misc_feature 11789..11876
/feature="Sequence confirmed by AC068749 sequenced by W1BR."
misc_feature 11877..11918
/feature="Sequence confirmed by AC068749 sequenced by W1BR."
repeat_region 12975..13178
/feature="MER58A repeat: matches 4. .224 of consensus"
repeat_region 13248..13525
/feature="AluX repeat: matches 1. .305 of consensus"
repeat_region 13540..14113
/feature="LNR34 repeat: matches 1. .699 of consensus"
repeat_region 15317..15612
/feature="AluSg repeat: matches 1. .236 of consensus"
repeat_region 16123..16298
/feature="MIR repeat: matches 79. .256 of consensus"
repeat_region 16561..16647
/feature="MIR repeat: matches 5. .82 of consensus"
repeat_region 16648..17011
/feature="MLT1A1 repeat: matches 1. .365 of consensus"
misc_feature 16860..17241
/feature="Single clone region. Assembly confirmed by restriction digest data."
repeat_region 17012..17068
/feature="MIR repeat: matches 82. .141 of consensus"
misc_feature 17012..17068
/feature="FSHB"
/feature="match: cDNAs: Em:AK017593 Em:AF257212 Em:AB029157 Em:AF106914 Em:X15493 Em:M36804 Em:M14853 Em:AF134151 Em:M35676 Em:AF008550 Em:E00632 match: ESTs: Em:BF704061 Em:BF702099 Em:AV754057 Em:AV755146 Em:AV754501 Em:AV753373"
/feature="product="dj562D20.1 (follicle stimulating hormone, beta polypeptide)"
evidence="not_experimental"
complement(17121..20538)
/feature="FSHB"
complement(17121)
/feature="FSHB"
complement(17219..17224)
/feature="FSHB"
complement(18136..18223)
/feature="L2 repeat: matches 2581. .2675 of consensus"
misc_feature 18217..18692
/feature="match: GSS: Em:AZ082576"
complement(18393..18595)
/feature="FSHB"
/feature="match: SRS: Em:G29876"
complement(join(18598..18828,20337..20495))
/feature="FSHB"
/feature="match: proteins: Tr:Q9QYB0 Tr:Q9JX69 Sw:Q60687 Sw:PI8427 Sw:O46430 Sw:P01225 Sw:P01228 Tr:Q9RTJ9 Sw:P04837 Sw:P01226"
/codon_start=1

repeat_region 20890..21053
/feature="AluB repeat: matches 134. .297 of consensus"
repeat_region 21813..22207
/feature="L2 repeat: matches 2244. .2628 of consensus"
repeat_region 22229..23211
/feature="L2 repeat: matches 1116. .2328 of consensus"
repeat_region 23212..23521
/feature="AluX repeat: matches 1. .310 of consensus"
repeat_region 23522..23788
/feature="L2 repeat: matches 890. .116 of consensus"
repeat_region 23800..24049
/feature="L25 copies 2 mer gg 55% conserved"
repeat_region 24113..24195
/feature="L2 repeat: matches 537. .619 of consensus"
misc_feature 24262..24541
/feature="Sequence from overlapping clone dj414F14 (AL391623). Assembly confirmed by restriction digest."
repeat_region 24628..24878
/feature="LTR repeat: matches 1. .256 of consensus"
misc_feature 24649..24918
/feature="Sequence from overlapping clone dj414F14 (AL391623). Assembly confirmed by restriction digest."
repeat_region 24954..25388
/feature="LTR repeat: matches 172. .691 of consensus"
repeat_region 25453..25750
/feature="AluX repeat: matches 1. .292 of consensus"
misc_feature 25586..26211
/feature="Sequence from overlapping clone dj414F14 (AL391623). Assembly confirmed by restriction digest."
repeat_region 25597..26093
/feature="match: GSS: Em:AQ440027"
repeat_region 25751..25806
/feature="MIR repeat: matches 118. .182 of consensus"
repeat_region 27093..27261
/feature="L2 repeat: matches 2383. .2545 of consensus"
repeat_region 27463..27484
/feature="L2 repeat: matches 11 copies 2 mer ta 100% conserved"
repeat_region 27582..27870
/feature="AluX repeat: matches 1. .288 of consensus"
repeat_region 28185..28497
/feature="MER33 repeat: matches 4. .324 of consensus"
repeat_region 29380..29596
/feature="MIR repeat: matches 28. .252 of consensus"
repeat_region 29895..30122
/feature="MIR repeat: matches 3. .232 of consensus"
misc_feature 30148..30186
/feature="Sequence from overlapping clone dj414F14 (AL391623). Assembly confirmed by restriction digest."
repeat_region 30420..30499
/feature="MLT1J repeat: matches 103. .183 of consensus"
repeat_region 30600..30755

```

Query Match 2.8%; Score 22; DB 9; Length 115865;
 Best Local Similarity 100.0%; Pred. NO. 2.4;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 tttagcagtgactctttctaa 350
 |||||
 Db 8342 TTGACAGTGACTCTTTCTAA 8321

Search completed: May 13, 2002, 23:34:41
 Job time: 7046 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 22:47:41 ; Search time 317.62 Seconds
(without alignments)
4216.338 Million cell updates/sec

Title: US-09-817-318-1
Perfect score: 780
Sequence: 1 aattactgttcttcttaag.....taagcgcaaaaaaaaaa 780

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780	100.0	780	AA19683	Human Mammary Gland
2	34	4.4	34	AA19706	Human MSG mam021 c
3	23	2.9	23	AA19703	Human MSG mam021 c
4	22	2.8	22	AA19704	Human MSG mam021 c
5	21	2.7	4607	AA198890	Human excretory re
6	21	2.7	4607	AA164069	Human bladder rela
7	21	2.7	61020	AA546788	Tumour suppressor
8	20	2.6	458	AAH69080	Human cervical can
9	20	2.6	490	AAH71943	Human cervical can

c	10	20	2.6	490	22	AAH72697	Human cervical can
c	11	20	2.6	550	22	AAH70743	Human cervical can
c	12	20	2.6	588	21	AAA16270	Human colon cancer
c	13	20	2.6	1092	23	ABL07199	Drosophila melanog
c	14	20	2.6	1113	21	AAAS1375	D. melanogaster dN
c	15	20	2.6	1124	22	AAAS3253	DNA encoding human
c	16	20	2.6	1352	19	AAV41918	Nucleotide sequenc
c	17	20	2.6	1518	20	AAV90449	Human secreted pro
c	18	20	2.6	1518	22	AAV59278	Human cDNA encodin
c	19	20	2.6	1518	24	ABA90947	Human polynucleoti
c	20	20	2.6	1877	21	AAV79705	Human secreted pro
c	21	20	2.6	3092	23	ABL07198	Drosophila melanog
c	22	20	2.6	5866	23	ABL07306	Drosophila melanog
c	23	20	2.6	7091	23	ABL07276	Drosophila melanog
c	24	20	2.6	7128	23	ABL07308	Drosophila melanog
c	25	20	2.6	7336	23	ABL07202	Drosophila melanog
c	26	20	2.6	9021	22	AAV46326	Tumour suppressor
c	27	20	2.6	9158	20	AAV03047	Human IL-1ra BAC c
c	28	20	2.6	11294	23	ABL04652	Drosophila melanog
c	29	20	2.6	21890	23	ABL15986	Drosophila melanog
c	30	20	2.6	22609	22	AAV35781	Human musculoskele
c	31	20	2.6	26928	20	AAV32184	Human prothrombin
c	32	20	2.6	32193	22	AAV35782	Human musculoskele
c	33	20	2.6	40862	24	ABL34072	Human immune syste
c	34	19	2.4	463	22	AAI89372	Human polynucleoti
c	35	19	2.4	481	20	AAV99731	Human adult testis
c	36	19	2.4	482	22	AAI94479	Human neuroblastom
c	37	19	2.4	498	21	AAV94508	Cat flea hindgut a
c	38	19	2.4	606	22	AAH09208	Human cDNA clone (
c	39	19	2.4	669	10	AAV90103	Human prealbumin c
c	40	19	2.4	789	22	AAH03248	Human cDNA clone (
c	41	19	2.4	891	22	AAI97716	Human neuroblastom
c	42	19	2.4	913	22	ABA20711	Human nervous syst
c	43	19	2.4	913	22	AAV36162	Human cardiovascular
c	44	19	2.4	914	22	ABA20710	Human nervous syst
c	45	19	2.4	914	22	AAV36161	Human cardiovascular

ALIGNMENTS

RESULT 1
AA19683
ID AA19683 standard; cDNA: 780 BP.
XX
XX
AC AA19683;
XX
DT 18-DEC-2001 (first entry)
XX
XX
DE Human Mammary Gland Cancer Specific Gene (MSG) mam021 cDNA.
XX
XX
KW Human; Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;
KW cancer; therapy; immune response; ss.
OS Homo sapiens.
PN WO200172780-A2.
XX
PD 04-OCT-2001.
XX
PF 26-MAR-2001; 2001WO-US09525.
XX
PR 27-MAR-2000; 2000US-192277P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Salceda S, Hu P, Recipon H, Caferkey R;
XX
DR WPI; 2001-616468/71.
XX
PT New isolated polynucleotide, mammary gland cancer specific gene (MSG),
PT useful for diagnosing, monitoring, staging, imaging and treating
PT mammary gland cancer

XX PS Claim 1; Page 87; 99pp; English.

CC The present sequence is human mammary gland cancer specific gene (MSG)

CC cDNA. MSG is useful for diagnosing, detecting, monitoring, staging,

CC prognosticating, imaging and treating mammary gland cancer in a patient

CC by determining the levels of MSG in cells, tissues or bodily fluids in a

CC patient and comparing the determined levels of MSG with levels of MSG

CC in cells, tissues or bodily fluids from a normal human control, where a

CC change in determined levels of MSG in the patient versus normal control

CC is associated with the presence of mammary gland cancer. MSG is used for

CC identifying potential therapeutic agents for use in imaging and treating

CC mammary gland cancer. MSG antibody conjugated to a cytotoxic agent is

CC useful for treating mammary gland cancer in a patient. MSG vaccine is

CC useful for inducing an immune response against a MSG protein and for

CC treating mammary gland cancer in a patient. MSG and its protein are

CC useful as diagnostic markers for mammary gland cancer and for diagnosis

CC and treatment of disorders of cells, tissues and organisms.

XX SQ Sequence 780 BP; 267 A; 115 C; 119 G; 279 T; 0 other;

Query Match 100.0%; Score 780; DB 22; Length 780;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aattactgttctcttaaaagtaaggcctcacccctactaaaatgtgatcaaaattttat 60

Db 1 aattactgttctcttaaaagtaaggcctcacccctactaaaatgtgatcaaaattttat 60

Qy 61 tatgaatagatgaaagctgtagctataaaattatgagagtaagttattttattttatc 120

Db 61 tatgaatagatgaaagctgtagctataaaattatgagagtaagttattttattttatc 120

Qy 121 caaatgtagtctaatagcataatagcaactcactcaaatcttagaataaaaaagaat 180

Db 121 caaatgtagtctaatagcataatagcaactcactcaaatcttagaataaaaaagaat 180

Qy 181 aaaaatgttaatttttggaggaatggttaattttttctacaaaattgtgacagcttt 240

Db 181 aaaaatgttaatttttggaggaatggttaattttttctacaaaattgtgacagcttt 240

Qy 241 acagaccttactctcaaatgactgaactgaacataacatacaaaaagggtcctgtttac 300

Db 241 acagaccttactctcaaatgactgaactgaacataacatacaaaaagggtcctgtttac 300

Qy 301 aaaaagatagtcgaagaacttcagaaatttttgacagtgactctttcttaaccctttaac 360

Db 301 aaaaagatagtcgaagaacttcagaaatttttgacagtgactctttcttaaccctttaac 360

Qy 361 caaatattatgaagtcctcctcctcctcctcctcctcctcctcctcctcctcctcct 420

Db 361 caaatattatgaagtcctcctcctcctcctcctcctcctcctcctcctcctcctcct 420

Qy 421 tctgtagttccttgcctataattgaagcagttctcctgaatacaccacaaactgattta 480

Db 421 tctgtagttccttgcctataattgaagcagttctcctgaatacaccacaaactgattta 480

Qy 481 tgaagcccatgcttttggaaaagatttgcaacttcgcttgcctcattttacattgact 540

Db 481 tgaagcccatgcttttggaaaagatttgcaacttcgcttgcctcattttacattgact 540

Qy 541 gtacttgcatgtattgctagattgactatcagttagtagacaatacaaaaagatttaga 600

Db 541 gtacttgcatgtattgctagattgactatcagttagtagacaatacaaaaagatttaga 600

Qy 601 taatggcagggaataatcagaagttactgtcaatacaaaagtattgtttttatgggtatt 660

Db 601 taatggcagggaataatcagaagttactgtcaatacaaaagtattgtttttatgggtatt 660

Qy 661 ttataggtgataaaattcattactagcaaatcttcaatcatcatgtttttattctctggtgt 720

Db 661 ttataggtgataaaattcattactagcaaatcttcaatcatcatgtttttattctctggtgt 720

Qy 721 aatatgtgactctggagactcaaatatttaattattgtttaaaggcaaaaaaaa 780

Db 721 aatatgtgactctggagactcaaatatttaattattgtttaaaggcaaaaaaaa 780

RESULT 2

AAAD19706/c

ID AAD19706 standard; DNA; 34 BP.

XX AC AAD19706;

XX AC

XX 18-DEC-2001 (first entry)

XX DE Human MSG mam021 cDNA amplifying Q-PCR probe #1.

XX DE

XX KW Human; Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;

XX KW cancer; therapy; immune response; PCR probe; ss.

XX OS Homo sapiens.

XX PN WC200172780-A2.

XX PD 04-OCT-2001.

XX PF 26-MAR-2001; 2001WO-US09525.

XX PR 27-MAR-2000; 2000US-192277P.

XX PA (DIAD-) DIADEXUS INC.

XX PI Salceda S, Hu P, Recipon H, Cafferkey R;

XX DR WPI; 2001-616468/71.

XX New isolated polynucleotide, mammary gland cancer specific gene (MSG),

XX useful for diagnosing, monitoring, staging, imaging and treating

XX mammary gland cancer

XX Example 3; Page 54; 99pp; English.

XX The present sequence is a PCR probe used for amplifying human mammary

XX gland cancer specific gene (MSG) cDNA. MSG is useful for diagnosing,

XX detecting, monitoring, staging, prognosticating, imaging and treating,

XX mammary gland cancer in a patient by determining the levels of MSG in

XX cells, tissues or bodily fluids in a patient and comparing the determined

XX levels of MSG with levels of MSG in cells, tissues or bodily fluids from

XX a normal human control, where a change in determined levels of MSG in the

XX patient versus normal control is associated with the presence of mammary

XX gland cancer. MSG is used for identifying potential therapeutic agents

XX for use in imaging and treating mammary gland cancer. MSG antibody

XX conjugated to a cytotoxic agent is useful for treating mammary gland

XX cancer in a patient. MSG vaccine is useful for inducing an immune

XX response against a MSG protein and for treating mammary gland cancer in

XX a patient. MSG and its protein are useful as diagnostic markers for

XX mammary gland cancer and for diagnosis and treatment of disorders of

XX cells, tissues and organisms.

XX Sequence 34 BP; 17 A; 6 C; 6 G; 5 T; 0 other;

Query Match

Best Local Similarity 4.4%; Score 34; DB 22; Length 34;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 414 gttttctctgagttccttgcctataattga 447

Db 34 GTTTTCTCTGTGAGTCTCTTTCCTATAATGA 1

RESULT 3

AAAD19703

ID AAD19703 standard; DNA; 23 BP.

```

XX AC AAD19703;
XX DT 18-DEC-2001 (first entry)
XX DE Human MSG mam021 cDNA amplifying Mam021 forward PCR primer.
XX DE
XX KW Human: Mammary Gland Cancer Specific Gene: MSG; cytostatic; vaccine;
XX KW cancer; therapy; immune response; PCR primer; ss.
XX OS Homo sapiens.
XX PN WO200172780-A2.
XX PD 04-OCT-2001.
XX PF 26-MAR-2001; 2001WO-US09525.
XX PR 27-MAR-2000; 2000US-192277P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Salceda S, Hu P, Recipon H, Caferkey R;
XX DR WPI; 2001-616468/71.
XX PT New isolated polynucleotide, mammary gland cancer specific gene (MSG),
XX PT useful for diagnosing, monitoring, staging, imaging and treating
XX PT mammary gland cancer -
XX PS Example 3; Page 54; 99pp; English.
XX CC The present sequence is a PCR primer used for amplifying human mammary
XX CC gland cancer specific gene (MSG) cDNA. MSG is useful for diagnosing,
XX CC detecting, monitoring, staging, prognosticating, imaging and treating
XX CC mammary gland cancer in a patient by determining the levels of MSG in
XX CC cells, tissues or bodily fluids in a patient and comparing the determined
XX CC levels of MSG with levels of MSG in cells, tissues or bodily fluids from
XX CC a normal human control, where a change in determined levels of MSG in the
XX CC patient versus normal control is associated with the presence of mammary
XX CC gland cancer. MSG is used for identifying potential therapeutic agents
XX CC for use in imaging and treating mammary gland cancer. MSG antibody
XX CC conjugated to a cytotoxic agent is useful for treating mammary gland
XX CC cancer in a patient. MSG vaccine is useful for inducing an immune
XX CC response against a MSG protein and for treating mammary gland cancer in
XX CC a patient. MSG and its protein are useful as diagnostic markers for
XX CC mammary gland cancer and for diagnosis and treatment of disorders of
XX CC cells, tissues and organisms.
XX SQ Sequence 23 BP; 4 A; 9 C; 1 G; 9 T; 0 other;

Query Match 2.9%; Score 23; DB 22; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 ccacgtcttcctttatccaact 400
Db 1 ccacgtcttcctttatccaact 23

RESULT 4
AAD19704/c
ID AAD19704 standard; DNA; 22 BP.
XX AC AAD19704;
XX DT 18-DEC-2001 (first entry)
XX DE Human MSG mam021 cDNA amplifying Mam021 reverse PCR primer.
XX DE
XX KW Human: Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;
XX KW cancer; therapy; immune response; PCR primer; ss.

```

```

XX OS Homo sapiens.
XX PN WO200172780-A2.
XX PD 04-OCT-2001.
XX PF 26-MAR-2001; 2001WO-US09525.
XX PR 27-MAR-2000; 2000US-192277P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Salceda S, Hu P, Recipon H, Caferkey R;
XX DR WPI; 2001-616468/71.
XX PT New isolated polynucleotide, mammary gland cancer specific gene (MSG),
XX PT useful for diagnosing, monitoring, staging, imaging and treating
XX PT mammary gland cancer -
XX PS Example 3; Page 54; 99pp; English.
XX CC The present sequence is a PCR primer used for amplifying human mammary
XX CC gland cancer specific gene (MSG) cDNA. MSG is useful for diagnosing,
XX CC detecting, monitoring, staging, prognosticating, imaging and treating
XX CC mammary gland cancer in a patient by determining the levels of MSG in
XX CC cells, tissues or bodily fluids in a patient and comparing the determined
XX CC levels of MSG with levels of MSG in cells, tissues or bodily fluids from
XX CC a normal human control, where a change in determined levels of MSG in the
XX CC patient versus normal control is associated with the presence of mammary
XX CC gland cancer. MSG is used for identifying potential therapeutic agents
XX CC for use in imaging and treating mammary gland cancer. MSG antibody
XX CC conjugated to a cytotoxic agent is useful for treating mammary gland
XX CC cancer in a patient. MSG vaccine is useful for inducing an immune
XX CC response against a MSG protein and for treating mammary gland cancer in
XX CC a patient. MSG and its protein are useful as diagnostic markers for
XX CC mammary gland cancer and for diagnosis and treatment of disorders of
XX CC cells, tissues and organisms.
XX SQ Sequence 22 BP; 5 A; 3 C; 7 G; 7 T; 0 other;

Query Match 2.8%; Score 22; DB 22; Length 22;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 448 agcagttctctgaaatcaccca 469
Db 22 AGCAGTTCTCTGAAATCACCCA 1

RESULT 5
AAI98890
ID AAI98890 standard; DNA; 4607 BP.
XX AC AAI98890;
XX DT 07-JAN-2002 (first entry)
XX DE Human excretory related polynucleotide SEQ ID NO 654.
XX KW Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
XX KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX KW antiparkinsonian; antisking; antianaemic; antiarthritic; cancer;
XX KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
XX KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX KW excretory system; ds.
XX OS Homo sapiens.

```

PN WO200155313-A2.
XX 02-AUG-2001.
PD 17-JAN-2001; 2001WO-US01323.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX
XX
PI

```

XX WPI: 2001-465569/50.
XX
XX Isolated nucleic acid molecule encoding excretory system antigen is
PT used in preventing, treating or ameliorating a medical condition -
XX
XX Example 2: SEQ ID NO 654; 574pp + Sequence Listing; English.
XX
XX The invention relates to novel excretory system related human
CC polynucleotides (AA198567-AA199503) and the encoded proteins
CC (AAW99594-AAW99913) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy, especially
CC disorders related to the excretory system. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4607 BP; 1225 A; 1140 C; 1222 G; 1020 T; 0 other;
SQ

Query Match      2.7%; Score 21; DB 22; Length 4607;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 aaaaaatgaataaaatgttaa 190
    |||||
Db 3656 aaaaaatgaataaaatgttaa 3676

RESULT 6
AA164069
ID AA164069 standard; cDNA; 4607 BP.
XX
AC AA164069;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human bladder related polynucleotide, SEQ ID NO: 102.
XX
KW Human; bladder antigen; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antinflammatory; antitumor; vulnary; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; gene therapy; cancer;
KW immune disorder; cardiovascular disorder; wound healing; infection;
KW neurological disease; ss.
XX
OS Homo sapiens.
XX
PN WO200159064-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01342.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR

```

PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246529.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251300.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-514652/56.
 Forty five bladder related polynucleotides, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases -
 PS Disclosure; SEQ ID NO 102; 482pp + sequence listing; English.
 XX
 CC The invention relates to forty five novel bladder related polynucleotides. The polynucleotides and the polypeptides that they encode are useful in the diagnosis, treatment and prevention of: CC cancer, particularly breast and ovarian cancer, and other cancers

CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital system; immune disorders such as Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular disorders such as myocardial ischaemias; wound healing; neurological diseases such as cerebral anoxia and epilepsy; and infectious diseases such as viral, bacterial, fungal and parasitic infections. Numerous examples of each type of disorder are given in the specification.
 CC The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The polynucleotides are useful for chromosome identification. They are also useful as probes for diagnosing or treating a disorder related to the female reproductive system, particularly breast and/or ovary cancer. The present sequence is a bladder related polynucleotide encompassed by the invention.
 CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 4607 BP; 1225 A; 1140 C; 1222 G; 1020 T; 0 other;

Query Match 2.7%; Score 21; DB 22; Length 4607;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 aaaaaatgaataaaatgttaa 190
 Db 3656 aaaaaatgaataaaatgttaa 3676
 |||||||
 RESULT 7
 AAS46788
 ID AAS46788 standard; DNA: 61020 BP.
 XX
 AC AAS46788;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #514.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP; cytosine methylation; ds.
 KW
 XX Homo sapiens.
 OS
 XX WO200168912-A2.
 PN
 XX 20-SEP-2001.
 PD
 XX 15-MAR-2001; 2001WO-EP02955.
 XX
 XX 15-MAR-2000; 2000DE-1013847.
 PR
 PR 06-APR-2000; 2000DE-1019058.
 PR
 PR 07-APR-2000; 2000DE-1019173.
 PR
 PR 30-JUN-2000; 2000DE-1032529.
 PR
 PR 01-SEP-2000; 2000DE-1043826.
 PR
 PR (EPIG-) EPIGENOMICS AG.
 PA
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor genes and oncogenes, useful in designing primers and probes for analysing diseases associated with cytosine methylation state e.g. cancer -
 PT
 XX Claim 1; SEQ ID No 514; 27pp; English.
 PS
 XX


```

Query Match          2.6%; Score 20; DB 22; Length 490;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
Db 31 TAAAGGCAAAAAA 12

RESULT 10
AAH72697/C
ID AAH72697 standard; cDNA; 490 BP.
XX
AC AAH72697;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 3971.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
PR 21-DEC-1999; 99US-0171350.
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 803; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 490 BP; 162 A; 94 C; 68 G; 163 T; 3 other;

Query Match          2.6%; Score 20; DB 22; Length 490;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
Db 31 TAAAGGCAAAAAA 12

RESULT 11
AAH70743/C
ID AAH70743 standard; cDNA; 550 BP.
XX
AC AAH70743;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 2017.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
PR 21-DEC-1999; 99US-0171350.
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 803; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 550 BP; 170 A; 109 C; 91 G; 180 T; 0 other;

Query Match          2.6%; Score 20; DB 22; Length 550;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780
Db 72 TAAAGGCAAAAAA 53

RESULT 12
AAH16270/C
ID AAH16270 standard; DNA; 588 BP.
XX
AC AAH16270;
XX
DT 14-JUN-2000 (first entry)
XX
DE Human colon cancer differentially expressed nucleotide sequence #275.
XX
KW Colon cancer; detect; differential expression; human; treatment;
KW detect mutation; non-invasive diagnostic method; ds.
XX
OS Homo sapiens.
XX
PN WO200012702-A2.
XX

```


PD 09-MAR-2000.
XX
PF 30-AUG-1999; 99WO-US19424.
XX
PR 31-AUG-1998; 98US-0098639.
PR 27-JAN-1999; 99US-0117393.
XX
PA (FARB) BAYER CORP.
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
PI Catino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
PI Schlegel R;
XX
DR WPI: 2000-256641/22.
XX
PT Novel nucleic acids and proteins for identifying therapeutic agents
PT useful for treating and diagnosing cancer, especially colon cancer
XX
PS Claim 16: Page 234; 345pp; English.
CC This sequence represents a human nucleotide sequence which is
CC differentially expressed in colon cancer cells compared to the expression
CC levels in normal cells. The nucleotide sequence can be used as a source
CC of primers and probes. The nucleotide sequence is useful for determining
CC the phenotype of a cell by detecting the differential expression of the
CC sequence relative to a normal cell. The probes derived from the sequence
CC can also be used to determine the phenotype of cells in a sample. Probes
CC and antibodies which hybridise to the nucleotide sequence can also be
CC used to determine the phenotype of a cell. The primers are useful for
CC detecting a mutation in a test nucleotide sequence and also for detecting
CC cancer, preferably colon cancer. Antibodies against the protein encoded
CC by the nucleotide sequence can also be used in a method to detect colon
CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
XX colon cancer at an early stage.
XX
SQ Sequence 588 BP; 187 A; 110 C; 91 G; 180 T; 20 other;

Query Match 2.6%; Score 20; DB 21; Length 588;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 761 taaaggcaaaaaa 780
Db 30 TAAAGGCAAAAAAAAAA 11

RESULT 13
ABL07199/c
ID ABL07199 standard; cDNA: 1092 BP.
XX
AC ABL07199;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16079.
XX
KW Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR P-PSDB; ABB63096.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1: SEQ ID NO 16079; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1092 BP; 324 A; 259 C; 248 G; 261 T; 0 other;

Query Match 2.6%; Score 20; DB 23; Length 1092;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 323 tgaattttgacagtgcac 342
Db 946 TGAATTTTGCACAGTGACTC 927

RESULT 14
AAA51375
ID AAA51375 standard; cDNA: 1113 BP.
XX
AC AAA51375;
XX
DT 26-SEP-2000 (first entry)
XX
DE D. melanogaster dNK, multisubstrate deoxyribonucleosidase cDNA.
XX
KW dNK; multisubstrate deoxyribonucleosidase; kinase; cytostatic;
KW virucide; ss.
XX
OS Drosophila melanogaster.
XX
FH Key Location/Qualifiers
FT CDS 113..865
FT /*tag= a
FT /product= Multisubstrate_deoxyribonucleosidase
XX
PN WO200036099-A1.
XX
PD 22-JUN-2000.
XX
PF 10-DEC-1999; 99WO-SE02314.
XX
PR 11-DEC-1998; 98SE-0004298.
XX
PA (KARL/) KARLSSON A.
XX
PI Karlsson A, Johansson M;
XX
DR WPI: 2000-431584/37.
DR P-PSDB; AAY96812.
XX
PT New nucleic acid encoding a multisubstrate deoxyribonucleosidase with
PT homology to a Drosophila melanogaster sequence for activating
PT anti-viral and anti-cancer drugs to treat viral infections and cancer
PT

XX Claim 1; Page 33; 38pp; English.

XX This cDNA encodes Drosophila melanogaster multibase substrate

CC deoxyribonucleoside (dNK). The multibase substrate deoxyribonucleoside

CC changes nucleoside analogue prodrugs into active anti-cancer and

CC anti-viral drugs by phosphorylation, allowing cancer and viral

CC infections to be treated. Drosophila melanogaster deoxyribonucleoside

CC kinase is (unlike human deoxyribonucleoside kinase) a multibase substrate

CC enzyme and it catalyzes phosphorylation of pyrimidine and purine

CC deoxyribonucleosides. The catalytic rates of deoxyribonucleoside and

CC nucleoside analogue phosphorylation are 10 to 100 fold higher than the

CC maximal catalytic rates reported for mammalian enzymes, making it more

XX useful for activating anti-viral and anti-cancer drugs.

XX Sequence 1113 BP; 324 A; 287 C; 272 G; 230 T; 0 other;

SQ

Query Match 2.6%; Score 20; DB 21; Length 1113;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780

Db 1087 taaaggcaaaaaa 1106

|||||

RESULT 15

AAS33253

ID AAS33253 standard; CDNA; 1124 BP.

XX AC AAS33253;

XX 04-DEC-2001 (first entry)

XX DNA encoding human secreted protein, Seq ID No 212.

XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;

KW rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;

KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;

KW cytosolic; Alzheimer's disease; Parkinson's disease; human; cancer;

KW multiple sclerosis; cancer; hyperproliferative disorder; infection;

KW Gaucher's disease; neurological disease; cerebrovascular disorder;

KW thrombosis; wound healing; ss.

XX Homo sapiens.

OS

XX WO200155326-A2.

PN

XX 02-AUG-2001.

PD

XX 17-JAN-2001; 2001WO-US01347.

XX

XX 31-JAN-2000; 2000US-0179065.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Barash SC, Ruben SM;

PI

XX WPI; 2001-451931/48.

DR

XX P-PSDB; AAU20544.

DR

XX New nucleic acids and polypeptides, useful for diagnosing, preventing

PT or treating medical conditions -

PT

XX Claim 1; SEQ ID No 212; 753pp; English.

PS

XX The invention relates to novel isolated nucleic acid molecules (I)

CC encoding human secreted proteins (II). (I) and (II) are used to prevent,

CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,

CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in

CC the prevention, treatment and diagnosis of diseases associated with

CC inappropriate expression of secreted proteins. (I) and complementary

CC

CC sequences may also be used as DNA probes in diagnostic assays (e.g.

CC polymerase chain reactions (PCR)) to detect and quantitate the presence

CC of similar nucleic acid sequences in samples, and so which patients may

CC be in need of restorative therapy. (II) may also be used as antigens in

CC the production of antibodies and in assays to identify modulators

CC (agonists and antagonists) of the expression and activity of the secreted

CC proteins. The anti-(II) antibodies and antagonists may also be used to

CC down regulate expression and activity of (II). The anti-(II) antibodies

CC may also be used as diagnostic agents for detecting the presence of (II)

CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The

CC disorders include for example: immune/autoimmune diseases (e.g. HIV

CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis

CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.

CC melanomas, neoplasms of the breast or liver, Sezary syndrome and

CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,

CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/

CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia,

CC angina and thrombosis), infections caused by bacteria, viruses and

CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),

CC agonists, antagonists and antibodies can also be used to promote wound

CC healing, maintain organs before transplantation, and support cell culture

CC of primary tissues. AAS33043-AAS33486 represent human secreted protein

CC coding sequences, PCR primers, and related sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification but was obtained in electronic format directly from WIPO

CC at: ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 1124 BP; 375 A; 228 C; 238 G; 283 T; 0 other;

Query Match 2.6%; Score 20; DB 22; Length 1124;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 taaaggcaaaaaa 780

Db 1099 taaaggcaaaaaa 1118

|||||

Search completed: May 14, 2002, 00:07:17

Job time: 4776 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 22:10:05 ; Search time 1946.33 Seconds

(without alignments)

5408.962 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgttctcttaaa.....taaggcaaaaaaaaaaaaaa 780

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	3.1	879	12 CNS04JBE	AL293315 Tetraodon
2	24	3.1	885	12 CNS04PKX	AL288474 Tetraodon
3	23	2.9	370	9 AI991804	AI991804 wf32606.x
4	23	2.9	481	9 AA970105	AA970105 op65f11.s
5	23	2.9	482	9 AW182002	AW182002 xj69f03.x
6	22	2.8	129	9 AI340638	AI340638 tb32c02.x
7	22	2.8	172	10 BF457830	BF457830 UI-M-BZ1-
8	22	2.8	333	10 BF611293	BF611293 dd78c05.y
9	22	2.8	349	9 AI706726	AI706726 UI-R-AA1-
10	22	2.8	374	9 AI927339	AI927339 wn49e04.x
11	22	2.8	393	10 BG664539	BG664539 DRABFB08
12	22	2.8	395	9 AA818458	AA818458 UI-R-A0-a
13	22	2.8	410	9 AA678990	AA678990 zf83e06.s
14	22	2.8	448	9 AA818109	AA818109 UI-R-A0-a
15	22	2.8	469	12 AZ913310	AZ913310 RPCI-24-1
16	22	2.8	483	9 AW046246	AW046246 UI-M-BH1-
17	22	2.8	497	9 AW529757	AW529757 UI-R-BU0-

18	2.8	520	10 BF162583	BF162583 601770225
19	2.8	532	12 AQ457564	AQ457564 HS.5097-B
20	2.8	545	12 BH120999	BH120999 RPCI-24-2
21	2.8	546	9 AA957399	AA957399 UI-R-E1-f
22	2.8	572	12 BH095483	BH095483 RPCI-24-2
23	2.7	227	10 BF847151	BF847151 CM3-EN004
24	2.7	297	10 BF847115	BF847115 CM3-EN004
25	2.7	328	9 AW511570	AW511570 xu60a08.x
26	2.7	328	12 AQ948067	AQ948067 Sheared D
27	2.7	329	9 AW469992	AW469992 xr27c12.x
28	2.7	348	10 BE950018	BE950018 UI-M-CD0-
29	2.7	387	10 BM148305	BM148305 TCAAPIT23
30	2.7	397	10 BG290862	BG290862 602387011
31	2.7	430	10 BF511890	BF511890 UI-H-B14-
32	2.7	448	10 BI158830	BI158830 602922030
33	2.7	492	10 BF395183	BF395183 UI-R-CM0-
34	2.7	495	10 BM145066	BM145066 TCAAPID14
35	2.7	678	12 AZ344764	AZ344764 IM0079A05
36	2.7	837	10 BF541441	BF541441 602069262
37	2.7	855	12 AZ189406	AZ189406 SP.1013-B
38	2.7	897	12 AZ530928	AZ530928 ENTBR71TR
39	2.7	948	12 CNS02799	AL184374 Tetraodon
40	2.7	1193	11 BC006015	BC006015 Homo sapi
41	2.6	77	9 AI345535	AI345535 tb69h12.x
42	2.6	118	10 BE464665	BE464665 hs86c05.x
43	2.6	147	10 BI274962	BI274962 UI-R-CX0-
44	2.6	148	10 BG403057	BG403057 602418871
45	2.6	165	10 BI403960	BI403960 MI-P-CPI-

ALIGNMENTS

RESULT	1
CNS04JBE	
LOCUS	
DEFINITION	CNS04JBE 879 bp DNA linear GSS 21-MAY-2000
ACCESSION	Tetraodon nigroviridis genome survey sequence T7 end of clone
VERSION	114M01 of library G from Tetraodon nigroviridis, genomic survey
KEYWORDS	sequence.
SOURCE	AL293315.1 GI:8031895
ORGANISM	GSS; genome survey sequence.
REFERENCE	Tetraodon nigroviridis
AUTHORS	Tetraodon nigroviridis
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
JOURNAL	1 (bases 1 to 879)
REFERENCE	Roest-Crollius.H., Jaillon.O., Dasilva.C., Fizes.C., Fisher.C., Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and Weissbach.J.
AUTHORS	Roest-Crollius.H., Jaillon.O., Dasilva.C., Fizes.C., Fisher.C., Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and Weissbach.J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
AUTHORS	2 (bases 1 to 879)
REFERENCE	Roest-Crollius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C., Bernot.A., Fizes.C., Wincker.P., Brottier.P., Quetier.F., Saurin.W. and Weissbach.J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 879)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES	Location/Qualifiers
source	1. .879
	/organism="Tetraodon nigroviridis"

```

/db_xref="taxon:99883"
/clone="114M01"
/clone_lib="G"
/note="Genoscope sequence ID : COBG114AG011P1-end : 77"
BASE COUNT 207 a 220 c 223 g 226 t 3 others
ORIGIN

Query Match
Best Local Similarity 3.1%; Score 24; DB 12; Length 879;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 aaatgaataaatagttaatttt 195
|||||
Db 851 AAAATGAATAAATGTTAATTTT 874

RESULT 2
CNS04FKX CNS04FKX 885 bp DNA linear GSS 21-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 106D12 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL288474
VERSION AL288474.1 GI:8027010
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 885)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 885)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 885)
Genoscope.
Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DDAR databases
JOURNAL This sequence is a single read and was generated as part of a large
AUTHORS scale clone-end sequencing project of the Tetraodon nigroviridis
COMMENT genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
Location/Qualifiers
1..885
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="106D12"
/clone_lib="G"
/note="Genoscope sequence ID : COBG106DB06SP1-end :
PUC-Ori"
BASE COUNT 230 a 249 c 193 g 210 t 3 others
ORIGIN

Query Match
Best Local Similarity 3.1%; Score 24; DB 12; Length 885;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 aaatgaataaatagttaatttt 195
|||||
Db* 26 AAAATGAATAAATGTTAATTTT 49

```

```

RESULT 3
AI991804/c AI991804 370 bp mRNA linear EST 08-MAR-2000
LOCUS wr32e06.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2489410 3',
DEFINITION mRNA sequence.
ACCESSION AI991804
VERSION AI991804.1 GI:5838709
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 370)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI_CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1360 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
Location/Qualifiers
1..370
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2489410"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
983608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 91 a 100 c 79 g 100 t
ORIGIN

Query Match
Best Local Similarity 2.9%; Score 23; DB 9; Length 370;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 758 gtgtaaaggcaaaaaa 780
|||||
Db 37 GTGTAAGGCAAAAAA 15

RESULT 4
AA970105/c AA970105 481 bp mRNA linear EST 07-JUL-1998
LOCUS OP65f11.s1 Soares,NFL_T-GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1581741 3' similar to TR:Q63742 Q63742 P34 PROTEIN. ; mRNA
sequence.
ACCESSION AA970105
VERSION AA970105.1 GI:3145618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 481)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 651 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 323.
Location/Qualifiers
1..481

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1581741"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI-CGAP.GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 110 a 131 c 116 g 124 t
ORIGIN

Query Match 2.9%; Score 23; DB 9; Length 481;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 758 gtgtaaaggcaaaaaa 780
|||||
Db 33 GTCTAAGGCAAAAAA 11

RESULT 5
AW182002/c
LOCUS xj69f03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2662493 3' similar to TR:Q63742 Q63742 P34 PROTEIN. ;, mRNA
sequence.
ACCESSION AW182002.1 GI:6450462
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 482)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 178.
Location/Qualifiers
1..482

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="IMAGE:2662493"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI-CGAP.GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 119 a 133 c 106 g 124 t
ORIGIN

Query Match 2.9%; Score 23; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. NO. 6.3e+02;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 758 gtgtaaaggcaaaaaa 780
|||||
Db 32 GTCTAAGGCAAAAAA 10

RESULT 6
AI340638
LOCUS tb32c02.x1 NCI_CGAP_HSC2 Homo sapiens cDNA clone IMAGE:2056034 3',
DEFINITION mRNA sequence.
ACCESSION AI340638
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 129)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 196 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..129

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2056034"
/clone_lib="NCI_CGAP_HSC2"
/tissue_type="stem cell 34+/38+"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: bone marrow; Vector: pAMP1; mRNA made from bone marrow, stem cells 34+/38+, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."
BASE COUNT 58 a 16 c 30 g 25 t

ORIGIN

Query Match 2.8%; Score 22; DB 9; Length 129;
 Best Local Similarity 100.0%; Pred. No. 3.4e+03;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780
 Db 83 TGTAAGGCACAAAAA 104

RESULT 7

BF457830/c
 LOCUS
 DEFINITION BF457830 172 bp mRNA linear EST 01-DEC-2000
 UI-M-BZ1-bky-e-11-0-UI.s1 NIH_BMAP_MHI2.S1 Mus musculus cDNA clone
 UI-M-BZ1-bky-e-11-0-UI 3', mRNA sequence.
 ACCESSION BF457830
 VERSION
 KEYWORDS
 SOURCE EST.
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 172)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT
 Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 hippocampus tissue cDNA library preparation. M.B. Soares Lab Clone
 distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
 GENETICS. It should be noted that Bento Soares is generating a
 small number of additional specialized non-redundant arrays of BMAP
 cDNAs whose availability will be considered under appropriate and
 limited collaborative arrangements. The following repetitive
 elements were found in this cDNA sequence: 1-76,
 >POLY_A#Simple_repeat
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source
 1..172
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BZ1-bky-e-11-0-UI"
 /clone_lib="NIH_BMAP_MHI2_S1"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p77T3D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; The
 NIH_BMAP_MHI2_S1 library is a subtracted library derived
 from NIH_BMAP_MHI2. NIH_BMAP_MHI2 is a library derived
 from mouse hippocampus tissue. For a detailed description
 of the library from which this clone was derived, please
 visit our web site at brainest.eng.uiowa.edu.
 TAG_LIB_NIH_BMAP_MHI2_S1
 TAG_TISSUE=hippocampus
 TAG_SEQ=NAGTC
 52 a 24 c 23 g 73 t

BASE COUNT
 ORIGIN

Query Match 2.8%; Score 22; DB 10; Length 172;
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 tagataaaaaatgaataaaat 185
 Db 32 TAGAATAAAAAATGAATAAAAT 11

RESULT 8

BF611293
 LOCUS
 DEFINITION BF611293 333 bp mRNA linear EST 14-DEC-2000
 dd78c05.y1 Wellcome CRC pcDNA1 egg xenopus laevis cDNA clone
 IMAGE:3430280 5', mRNA sequence.
 ACCESSION BF611293
 VERSION
 KEYWORDS
 SOURCE EST.
 African clawed frog.
 ORGANISM
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 333)
 REFERENCE
 AUTHORS Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
 Martin, J., Willie, T., Underwood, K., Theising, B., Bowers, Y., Person,
 B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
 Waterston, R. and Wilson, R.
 TITLE WashU Xenopus EST project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Other_ESTs: dd78c05.x2
 Contact: Sandy Clifton, Ph.D.
 WashU Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
 Gurdon, (Wellcome/CRC Institute). DNA sequencing by: Washington
 University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -40RP from Gibco.

FEATURES

source
 1..333
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3430280"
 /clone_lib="Wellcome CRC pcDNA1 egg"
 /tissue_type="egg"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pcDNA1; Site 1: NotI; Site 2: EcoRI; cDNAs
 were oligo-dT primed and directionally cloned. Library was
 constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
 Gurdon (Wellcome/CRC Institute)."
 BASE COUNT 125 a 64 c 57 g 87 t
 ORIGIN

Query Match 2.8%; Score 22; DB 10; Length 333;
 Best Local Similarity 100.0%; Pred. No. 1.7e+03;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaagggcaaaaaa 780
 Db 307 TGTAAGGCACAAAAA 328

RESULT 9
 A1706726/c
 LOCUS

A1706726 349 bp mRNA linear EST 03-JUN-1999

DEFINITION UI-R-AAI-aab-a-12-0-UI.s1 UI-R-AAI Rattus norvegicus cDNA clone
ACCESSION UI-R-AAI-aab-a-12-0-UI 3', mRNA sequence.
VERSION A1706726
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 349)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers	source
1..349	/organism="Rattus norvegicus"
	/strain="Sprague-Dawley"
	/db_xref="taxon:10116"
	/clone="UI-R-AAI-aab-a-12-0-UI"
	/clone_lib="UI-R-AAI"
	/dev_stage="adult"
	/lab_host="DH10B (Life Technologies)"
	/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-AAI library is a normalized library constructed from 16.5 dpc rat atrium. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Jim Lin, Department of Biology, University of Iowa."
BASE COUNT	99 a 71 c 60 g 119 t
ORIGIN	TAG_SEQ=None found"

Query Match 2.8%; Score 22; DB 9; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaaggcacaagcaaaaaaa 780
|||||
Db 26 TGTAAGGCCAAAAA 5

RESULT 10
LOCUS AI927339/c
DEFINITION wn49e04.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2448798 3',
mRNA sequence.
ACCESSION AI927339
VERSION AI927339.1 GI:5663303
KEYWORDS EST.
SOURCE human.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

DEFINITION UI-R-AAI-aab-a-12-0-UI.s1 UI-R-AAI Rattus norvegicus cDNA clone
ACCESSION UI-R-AAI-aab-a-12-0-UI 3', mRNA sequence.
VERSION A1706726.1 GI:4994626
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 349)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers	source
1..349	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:2448798"
	/clone_lib="NCI_CGAP_Lu19"
	/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
	/dev_stage="adult"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	137 a 45 c 55 g 137 t
ORIGIN	

Query Match 2.8%; Score 22; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tgtaaggccaaaagcaaaaaaa 780
|||||
Db 28 TGTAAGGCCAAAAA 7

RESULT 11
LOCUS BG664539
DEFINITION DRABFB08 Rat DRG Library Rattus norvegicus cDNA clone DRABFB08 5',
mRNA sequence.
ACCESSION BG664539
VERSION BG664539.1 GI:13886461
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 393)
AUTHORS Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,
Guo,C., Yan.Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen.Z. and
Zhang,X.
TITLE Distinct gene expression profiles of rat dorsal root ganglion
induced by peripheral nerve axotomy
JOURNAL Unpublished (2001)
COMMENT Contact: Zhang Xu
Laboratory of Sensory System

Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn

This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)

PCR PRIMERS
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES

source
Location/Qualifiers
1. .393
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="DRABF08"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
BASE COUNT 129 a 65 c 78 g 121 t
ORIGIN

Query Match 2.8%; Score 22; DB 10; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaaggcaaaaaa 780
Db 365 TGTAAAGGCAAAAAA 386

RESULT 12

AA818458/c
LOCUS
DEFINITION
UT-R-A0-at-b-01-0-UI.s1 395 bp mRNA linear EST 03-JUL-1999
UT-R-A0-at-b-01-0-UI 3', similar to dbj|AB003515|AB003515 Rat mRNA
for GEF-2, complete cds, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

AA818458
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 395)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
On Feb 17, 1998 this sequence version replaced gi:2889197.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult Muscle library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics This clone is also available through the I.M.A.G.E.
Consortium at LNL (info@image.llnl.gov). IMAGE ID=1768015
Seq primer: M13 Forward
POLYA-No.

FEATURES

source
Location/Qualifiers
1. .395
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-A0-at-b-01-0-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; this library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
, liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."
BASE COUNT 123 a 78 c 65 g 129 t
ORIGIN

Query Match 2.8%; Score 22; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaaggcaaaaaa 780
Db 26 TGTAAAGGCAAAAAA 5

RESULT 13

AA678990
LOCUS
DEFINITION
zf83e06.s1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone
IMAGE:383554 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA678990
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 410)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE
JOURNAL
COMMENT

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 398.

FEATURES

source

Location/Qualifiers
1. .410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:383554"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTTTTTTTTT 3']
, double-stranded cDNA was size selected, ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M. Fatima Bonaldo. "

BASE COUNT 146 a 48 c 80 g 136 t

ORIGIN

Query Match 2.8%; Score 22; DB 9; Length 410;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaggcaaaaaa 780
|||||
Db 116 TGTAAAGGCAAAAAA 137

RESULT 14

AA818109/c

LOCUS

DEFINITION AA818109 448 bp mRNA linear EST 27-JAN-1999
UI-R-A0-am-f-09-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone

ACCESSION UI-R-A0-am-f-09-0-UI 3', mRNA sequence.

VERSION AA818109

KEYWORDS AA818109.1 GI:4198624

SOURCE EST.

ORGANISM Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 448)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

On Mar 9, 1998 this sequence version replaced gi:2946805.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dr track served to identify it as a clone from the normalized

adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo,

Ph.D. Clone distribution: clones will be available through Research

Genetics

Seq primer: M13 Forward.

Location/Qualifiers

1. 448

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-A0-am-f-09-0-UI"

/clone_lib="UI-R-A0"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73B-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; This library

consists of a mixture of individually tagged normalized

libraries constructed from rat placenta, adult lung, brain

, liver, kidney, heart, spleen, ovary, and muscle. The tag

is a string of 3-5 nucleotides present between the Not I

site and the oligo-dr track which allows identification of

the library of origin of a clone within the mixture."

BASE COUNT 149 a 94 c 81 g 124 t

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 22; DB 9; Length 448;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaggcaaaaaa 780

|||||

Db 26 TGTAAAGGCAAAAAA 5

RESULT 15

AZ913310

LOCUS

DEFINITION AZ913310

RPCI-24-167H1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-167H1,

DNA sequence.

ACCESSION AZ913310

VERSION AZ913310.1 GI:13232255

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 469)

AUTHORS

Zhao, S., Nierman, W., Malek, J., Shvartsbeyn, A., Gelegeorgis, E.,

Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gelegeorgis, E.,

Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Other GSSs: RPCI-24-167H1.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html;

Plate: 167 row: H column: 1

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 469

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-167H1"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 147 a 106 c 106 g 110 t

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 22; DB 12; Length 469;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 759 tgtaaggcaaaaaa 780

|||||

Db 447 TGTAAAGGCAAAAAA 468

Search completed: May 14, 2002, 00:01:29

Job time: 6684 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 21:33:45 ; Search time 2160.77 seconds
(without alignments)
7554.112 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgtgtcttctaaag.....taagcgcaaaaaaaaaa 780

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_on.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

ALIGNMENTS

RESULT 1

AX260297

LOCUS

DEFINITION

AX260297

ACCESSION

AX260297.1

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

BASE COUNT

ORIGIN

AX260297 100.0 780 780 bp DNA linear PAT 26-OCT-2001

Sequence 1 from Patent WO0172780.

AX260297

AX260297.1 GI:16509264

human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (sites)

Salceda S., Hu, P., Recipon, H. and Cafferkey, R.

Compositions and methods of diagnosing, monitoring, staging,

imaging and treating mammary gland cancer

Patent: WO 0172780-A 1 04-OCT-2001;

diadex, Inc. (US)

Location/Qualifiers

1..780

/organism="Homo sapiens"

/db_xref="taxon:9606"

267 a 115 c 119 g 279 t

Query Match		100.0%;	Score 780;	DB 6;	Length 780;				
Best Local Similarity		100.0%;	Pred. No. 2.3e-123;						
Matches 780;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;				
QY	1	aattactgtctctcttaaaagtagagccttacaccctactaaaaatgtgatcaaaattttat	60						
Db	1	AATTACTGTGTCCTCTTAAAGTAGAGGCTTACACCCCTACTAAAAATGTGATCAAAATTTTAT	60						
QY	61	tatgaatagatgaagaagctgtactataaattatgagagtaagtattttattattatc	120						
Db	61	TATGAATAGATGAAGAAGCTGTAGCTATAAATTATGAGAGTAAGCTTTATTTTATATTATC	120						
QY	121	caaatgtactctctcacaattgacattgacattcactaaatctagaataaaaaatgaat	180						
Db	121	CAAAATGTAGTTCAATATAGCATTAATAGCAACTTCACATAATCTTAGAATATAAANAATGAAT	180						
QY	181	aaaatgttaatttttggaggaatggttaatttttctacaaaattgtgtgacagcttt	240						
Db	181	AAAATGTTAATTTTGGAGGAATGGTTAATTTTCTACAAAATGTGTGACAGCTTT	240						
QY	241	acagaccttactctcacaaattgactgaacattatgagagtaagtattttattattac	300						
Db	241	ACAGACCTTACTCTTCAAAATGACTTGAACATTAACATCACAAGAGGGTCTGTGTTAC	300						
QY	301	aaagaatagtcgaagctcactgaattttgacagtgactcttttctaaccttttaac	360						
Db	301	AAAGAATAGTCGAAGAACTTCATGAATTTTGGACGTGACTCTTTCTAACCCCTTAATC	360						
QY	361	caaatattatttaagtccatcgctcttcttccaaactcattgtttaaactagtttct	420						
Db	361	CAAAATATATTTAAGTGTCCATCGCTTCCCTTTATCCAACCTCAATTTGTTAACTAGTTTCT	420						
QY	421	tcgtgaattcccttgcctataattgaagcagtgctctctgaaatcaccacaaactgattta	480						
Db	421	TCGTGAGTTTCCCTTGGCTATAATTGAAGCAGTCTCTGTAATACCAACCACTGATTTTA	480						
QY	481	tgaagcccatgccttttggaaagatttgcaacttcggtcttgcaatctattacattgact	540						
Db	481	TGAAGCCCATGCTTTTGGAAAGATTTGCACTTCGGCTTTGCAATCTATTTTACATTGACT	540						
QY	541	gtactgtcattgtattgtctagattgtgactatcagtttaggacaaatcaaaaagatttaga	600						
Db	541	GTACTGTCAATTGTATTGCTAGATTGTGACTATCATGTTAGGACAAATCAAAAGATATTAGA	600						
QY	601	taatggcaggataataatcagaagttactgtcaatacaaaagttattgtttatgggtatt	660						
Db	601	TAATGGCAGGGATAAATCAATCAGAAGTTACTGTCAATAACAAGTTATGTTTATGGGTATT	660						
QY	661	tatagtgataaattcattactgagcaatttcatactatgtttttaaattctctggtgtg	720						
Db	661	TTATAGTGATAAATTCATTACTGAGCAATTTTCATATCATGTTTAAATTCCTCCTGGTGT	720						
QY	721	aatatggctgactctggagactcaaatataattatgttggtgaagggcaaaaaa	780						
Db	721	AATATGGTGACTCTGGAGACTCAAAATATTAAATATTGCTGTAAGGCAAAAAA	780						
RESULT 2									
AC079414/c									
LOCUS									
DEFINITION Homo sapiens chromosome 16 clone RP11-358L22, complete sequence.									
AC079414									
AC079414									
AC079414.5									
GI:18057080									
HTG.									
human.									
Homo sapiens									
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
1 (bases 1 to 167021)									
DOE Joint Genome Institute.									
Sequencing of Human Chromosome 16									
REFERENCE									
AUTHORS									
TITLE									

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 167021)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	3 (bases 1 to 167021)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On Jan 4, 2002 this sequence version replaced gi:17976466. Sequence Quality Assessment: This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file. ----- Sequence Quality Assessment: This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file. -----
FEATURES	Location/Qualifiers
source	1. 167021 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /clone="RP11-358L22"
BASE COUNT	49239 a 36098 c 35465 g 46219 t
ORIGIN	
Query Match	92.9%; Score 724.6; DB 9; Length 167021;
Best Local Similarity	98.3%; Pred. No. 1.6e-114;
Matches	756; Conservative 0; Mismatches 4; Indels 9; Gaps 2;
QY	1 aattactgtctctttaaagtagagccttacaccctactaaaaatgtgatcaaaattttat 60
Db	94402 AATTACTGTGTCCTCTTAAAGTAGAGGCTTACACCTTACTAAATGTGATCAAAATTTTAT 94343
QY	61 tatgaatagatgaagaagctgtactataaaattatgagagtaagtattttattattatc 120
Db	94342 TATGAATAGATGAAGAAGCTGTAGCTATAAATATGACAGTAAGTTTATTTTATATTATC 94283
QY	121 caaatgtagtctcaatagcataatagcaacttcactaaatcttagaataaaaaatgaat 180
Db	94282 CAAATGTAGTT-----CATATAGCAACTTCACATAAATCTTAGAATAAANAATGAAT 94231
QY	181 aaaatgttaatttttggaggaatggttaatttttctacaaaattgtgtgacagcttt 240
Db	94230 AAAATGTTAATTTTGGAGGAATATGTTAATTTTCTACAAAATGTGTGACAGCTTT 94171
QY	241 acagaccttactctcacaattgacattgacattcactaaatcacaagaagggctctgttac 300
Db	94170 ACAGACCTTACTCTTCAAAATGTGACTTTGAACATTAACATCACAAGAGGGTCTGTGAC 94111
QY	301 aaaagaatagtcgaagaactcactgaatttttggacagtgactcttttctaaccttttaac 360
Db	94110 AAAAGAATAGTCGAAGAACTTCATGAATTTTGGACAGTACTCTTTCTAACCCCTTAATC 94051
QY	361 caaatattatttaagtgtccatcgctcttcttccatccactcaattgtttaactagtttct 420
Db	94050 CAAATATATTTAAGTGTCATCGTCGTCCTTTATCCAACTCATTTGTTAACTAGTTTCT 93991

Qy 421 tctgtgagttctcttgcctataaattgaagcaggttctctgaaatccaccacaaactgatatta 480
|||||
Db 9390 TCTGTGAGTCTCTTGGCTCTAATAATTGAAGCAGTCTCTGAAATCACCACCAACTGATTTTA 93931
|||||
Qy 481 tgaagcccatgcttttggaaagatttgcaactcgggttgcattcattacattgact 540
|||||
Db 9390 TGAAGCCCATGCTTTTGGAAAGATTTGCACCTCGGCTTTGCAATCTATTACATTGACT 93871
|||||
Qy 541 gtacttgatgtattgttagatgttgactatcagttaggacaatacaaaagatataga 600
|||||
Db 93870 GTACTTGCAATTGTATTGCTAGATGTGACTATCATGTTAGGACAATCAAAAATATATATT-GA 93812
|||||
Qy 601 taatgggcagggataaatacagaagttaactgtcaatacaaaagttaattgattggatt 660
|||||
Db 93811 TAATGGGCAGGGATAAATCAGAAGTTACTGTCAATAACAAGTTATGTTTATGGGTATT 93752
|||||
Qy 661 ttataggtgataaatttcattactgagcaattcatatcatatgcttttaattctctcgtgtgt 720
|||||
Db 93751 TTATAGGTGATAAATTCATTACTAGCAATTTTCATATCATGTTTAAATCTCTCGTGTGT 93692
|||||
Qy 721 aataggtgactctggaactcaaatattaaattgggtgtaaaaggcaa 769
|||||
Db 93691 AATAGGTGACTCTGGAGACTCAAAATATTAATATTTGCTGTAAGACAA 93643
|||||

RESULT 3

AC009979/c

LOCUS

DEFINITION Homo sapiens chromosome 12 clone RP11-324M3, WORKING DRAFT

SEQUENCE, 3 unordered pieces.

AC009979

VERSION AC009979.15 GI:15628175

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 192126)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,

Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Hollway,C.,

Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louisaged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

Mei,G., Metker,M., Miner,G., Miner,Z., Mitchell,T., Monabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S.,

Ogih,M., Okuwona,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,

Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (09-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 17, 2001 this sequence version replaced gi:15529027.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMJ
Center clone name: RP11-324M3
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 15% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 193928 bases at least Q40
Consensus quality: 195261 bases at least Q30
Consensus quality: 196330 bases at least Q20
Estimated insert size: 192447; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 3 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

* 1 149093: contig of 149093 bp in length
* 149094 149193: gap of unknown length
* 149194 189670: contig of 40477 bp in length
* 189671 189770: gap of unknown length
* 189771 192126: contig of 2356 bp in length.

FEATURES

Source

1.192126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-324M3"
BASE COUNT 52796 a 41379 c 41147 g 56577 t 227 others
ORIGIN

Query Match

Best Local Similarity 48.7%; Pred. No. 0.00047;

Matches 201; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

Qy 135 aatagcataatagcaacttcactaatcttagaataaaaaataaataatcttaatttt 194

|||||

Db 50638 AGTTGTAAGTCAACAATGTTCTGTGATTCTTAGAATCAACAAGTTCAAATCAAGTTTGG 50579

|||||

Qy 195 ttgagggaatggttaatttttctacaaaattgtgacagctttacagaccttact 254

|||||

Db 50578 CAAGTCTCTGAAACTATATATTTACATAAATTTATATGATACCAATGATGATGTTCC 50519

|||||

Qy 255 tcacatgacttgacacattaacatcacaaagagggtcctgtttacaaaagaatagtc 314

|||||

/note="contains similarity to Pfam family PF00078 (Reverse transcriptase (RNA-dependent DNA polymerase)), score=57.5, E=2.9e-13, N=1"

/codon_start=1

/product="Hypothetical protein Y76B12C.5"

/protein_id="AAF36061.1"

/db_xref="GI:7105675"

/translation="MLPFLAASPLGSDVRAPAIWLGLVSPSPRTPIITKNSKTKRKRTD GSNTROSSLGPSPSSPLRSAGGOLRKTTRSELAKSHFHLGSMNAQSLGLPRLASGVY LKNSCHVNLQETSPRSFMSASIDQPEQLLFSPKTSKDLGLGFFYKCKDQLEVINVS WIGVGLASLEFERKVKIMVINYAPVCKSRSENDAKRFERLUAETQURKQSKGPI VUGGDNAAATSDNDELAPWICGNVFGNSNNHDFEFNFSLAFQNLRSFRKRLA SWRTGAGKVGSTEDPFI SNRDLVEDVSTFNNLHNSDLRLIIRSWYISVKSERD HAFKSRPSTGTERDWTLYEADIKDLRHATLGSYDSFVKTLRQGLVPLPAIKPHKF SORTIMILORROVLSSVPDIAKRLKSOEARKSITEDITOKYDLSLEACRKRSSK ILNKSIPAEIPLYLWPDGSLTANRIKIRKIVESHFKTLYEAPTSPPSPRPNSNP PNSDPVAILKCBVRLEIKLKTSAPGLNDVDAAMLKNGDVTVDVSATFALNDILHH NKVDPLKADVLKIPKAKATKIDFPIISLPLILSKMFSSTLRRLTFLTESYLDE SOTKVPKRCRCADNITQSITMLIEKNEFQLPLLLFIDYQAFDKISHSVSSLLER AGADPMRKMIOELMEGGQAEITVHOKLKNLCTGIRQGDSPALFSAALQAILTD CDELAGVIGVSGRHILRLFEADVVILICSTPEEVOERLEILDRTSSNGLKINOSK TVLLKNKFCRSODILFNGSPIIPVPCGVILGRWIDISGIDEISRIRAGWALVGI KEVLIRMPKNERILLKQNVLPALLYASETWCNAGSTLRLKRTIVTGLIDAAEIRGNW FNLERVLLAKQSRFAGHILRRDPF"

16413..21859

/gene="Y76B12C.6"

join(16413..16851,18265..18363,19373..19580,20673..21468, 21659..21859)

/gene="Y76B12C.6"

/note="coded for by the following C. elegans cDNAs:

yk20166.5, yk459e5.5"

/codon_start=1

/product="Hypothetical protein Y76B12C.6"

/protein_id="AAF36066.1"

/db_xref="GI:7105680"

/translation="MGLEYLKSMLKERAFIIFKRILEYFIVHRLNHTGYKTFESAAGA OIMRRLIMYVAQOILIDFVMPVYRQAFYFRELQRLKILNPDWAMRYMTEFAQ REAFYREARISITREGLQELQVLPDEHPNSFPPTPEPDPRIETKTHFF AVFAGTPKPTNEOMASQLTUNKPTNOMSASRLDAAIILAMNTNLPPIAGDO PVDTLRLNLSMMNVALTNFQPTKNOEOREAEFPDPMRAMAIPHDKDLVLELPQSDA CRNTYSPRVGHMYGVPVPGSSAADALRVQAEHROPGEYQKQSDTYIGLTKITFE LNSADATGKVLQYQVDHTTSPSTAGFMULFDPIVPAAVOQASAPPAPPOLOGIDIA TLOSFGAPGAAPQPIPELDPGPPRLGPTPLVPIPNVETESSLGVAKRGARDG LGSGLSLDLARDVLDVAPPLPGGPIHGAQOQOQAVRKRDPDHOQOQSEENRRKM TVEGARSLTFCEIKPKNFWLELLKKKESRNSSEFQKNWAKFEFLRKLKKNKFEK KVEKLDFRGILRR"

23205..38006

/gene="Y76B12C.2"

join(23205..23327,23762..24752,26575..27035,31016..31424, 32544..32763,34005..34302,35754..35929,36655..36748, 37743..38006)

/gene="Y76B12C.2"

/note="contains similarity to xeroderma pigmentosum group C proteins; coded for by the following C. elegans cDNAs: yk116e7.3, yk116e7.5"

/codon_start=1

/product="Hypothetical protein Y76B12C.2"

/protein_id="AAF36063.1"

/db_xref="GI:7105677"

/translation="METRRSSRLQOQTTSDLNAREEPQSEPPVVRKARGAKNEAPS TAPMPKTKTKTPKQSGSPIAEILDGMELENEAEIIVQKSKNGRKLKVID ENLRISAENGSSKSNLEENRMEIDKPGNLAKKSKNGSRVVKSDKSNLVQSPVK STTNGSKVAIIEDDPIRAENGVKSKDEKPDFSAQNGSKLAONAPNRLSRPRSVT TAKKVSYVSDQOLELSSSSSELESSEDETEIRPTGSKIAKKEKSPKISSESS SEPPDDESASESDPSIPGSEPRKRRKIOKRLSTLSSGGATTKDLHWPCKSKASIA RTNPGSKNAARKPAKSLRWAKQKQPPQWKNLKDYEITDRKLAKGERMLEYRHV AIYAVARGKIEETTYDEYKLEHMKKAYFAGRSLLDAAVLDPVEPEKSNKYKSEK NDEKNTAGSSSEDEWEHEFOPPIIDNIEVSDHEGGDDGDEEVVYKDWAIYL ROEINKIREMENTHKLFLCRLGLLELTRLVNVYRAIPRRWDKTOQKLONELSKF RELRSRTTTPGKSLEIUEQKSGSKPAKQKSEKKAARKKVVEERYNWEYVWQF REKRWICVPLHKSVDLEIUEHSASPSYFAIDNKGICVDSYKQDFR RRTNPKVAVMTLFLPPFAANSRKKWEMQMRDLVKRPLPTVMSEYKHNHLYALEK DLLKFEAIVPPATQKPLQGIQRGHNVPRSTVTLQENNNWLKARSVKIGEPYKIV

KARDPRIIPVEDREDKFLDYGYWOTEKYRRPPLKNGKIPHNEYGNVYFMNEMCPLOD CTYLLKSLGLVQISRKLKQKQIPAVVWGAFDGGFTHPVIDGAIVLEKDAIDRFNAMEKL

Query Match 8.9%; Score 69.8; DB 3; Length 76485;
 Best Local Similarity 49.7%; Pred. No. 0.0028;
 Matches 261; Conservative 0; Mismatches 257; Indels 7; Gaps 3;

Qy 58 tattatgaatagatgaaagcgtgtagctataaattatgagataagatttatattatttt 117
 Db 35805 TATTTTCATTAACATGTACAGCTTGCCNATCTCGTGTGCGGAATTTTACCTGAAAAA 35746

Qy 118 atccaaatgtagtcataatagcataatgacacttcacataatctctagaataaaaaatg 177
 Db 35745 AAGGAAAT-TTGAAATTAATAATAAAAAATTTTTTTTTTTTTTGAATAATAAATTTT 35687

Qy 178 aataaaatgtaattttttggaggaatggtaatttttctacaaaaatctgtgacagc 237
 Db 35686 GTTTTGAATAATTAATTTTTTTGAAAAATATTTTTTTTTTGAATAATAAATTTT 35627

Qy 238 -tttacagacaccttactcttcacaaatgacttgacattgaacatcaacaaagagggcctgt 296
 Db 35626 TTTTGTGAAAAATATTTTTTTTGAATAATAAATTTTGTGAAAAATTAATTTTTTTT 35567

Qy 297 ttcaaaagaatagtcagaagaacttcataatgtttgacagtgactctttctaaacccctt 356
 Db 35566 TGAATAATATTTTTTTTGAATAATATTTTATTTTGAATAATATTTTTTTTGAATAATTT 35507

Qy 357 aatccaaatataatttaagtgtccactgcctctcttatacaactcattgttaactagt 416
 Db 35506 TTTTGTGAAAAATATTTTTTTTGAATAATAAATTTTTTTTGAATAATATTTTTTTTGAATAAT 35447

Qy 417 tctctgtgagctccttgcctataatgaagcagtgactctctgaaatcaccacaaactgat 476
 Db 35446 TTTTGTGAAAAATTTTTTTTTTGAATAATAAATTTTCAATCTCTTGAATAAT----AAATTTT 35392

Qy 477 tttatgaagcccatgcttttggaaagattgcacttcggttcgctcaactctattacatt 536
 Db 35391 TTTTGTGAAAAATTTTTTTTTTGAATAATTTTTTTTTTGAATAATAAATTTTATTTTAAAAATA 35332

Qy 537 gactgtacttcgattgattgcttagatgttgactatcagtttagga 581
 Db 35331 AATTTTTTTTTTAATAAATTTTTTTTTTTTGTGAGATTCAAAATTCGA 35287

RESULT 5

AC024872/6 76485 bp DNA linear INV 01-MAR-2000

LOCUS Caenorhabditis elegans clone Y76B12C, complete sequence.

DEFINITION AC024872

ACCESSION AC024872

VERSION AC024872.1 GI:7140435

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

REFERENCE 1 (bases 1 to 76485)
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

AUTHORS Waterston,R.H.

TITLE The sequence of Caenorhabditis elegans clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 76485)
 Waterston,R.H.

AUTHORS Direct Submission

TITLE Submitted (01-MAR-2000) Genome Sequencing Center, Washington

JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

FEATURES

source 1..76485
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="Y76B12C"

BASE COUNT 25107 a 13479 c 13044 g 24855 t

ORIGIN

Sequencing vector:	M13; 46%
Chemistry:	Dye-terminator; plasmid; 54%
Chemistry:	Dye-terminator ET; 0% of reads
Chemistry:	Dye-terminator Big Dye; 100% of reads
Assembly program:	Phrap; version 0.990319
Consensus quality:	189918 bases at least Q40
Consensus quality:	192873 bases at least Q30
Consensus quality:	194155 bases at least Q20
<p>*** NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.</p>	
1	1156: contig of 1156 bp in length
1157	1256: gap of unknown length
1257	2450: contig of 1194 bp in length
2451	2550: gap of unknown length
2551	3768: contig of 1218 bp in length
3769	3868: gap of unknown length
3869	5007: contig of 1139 bp in length
5008	5107: gap of unknown length
5108	7155: contig of 2048 bp in length
7156	7255: gap of unknown length
7256	8408: contig of 1153 bp in length
8409	8509: gap of unknown length
8509	12851: contig of 4342 bp in length
12851	12950: gap of unknown length
12951	19909: contig of 6959 bp in length
19910	20009: gap of unknown length
20010	27734: contig of 7725 bp in length
27735	27834: gap of unknown length
27835	48551: contig of 20717 bp in length
48552	48651: gap of unknown length
48652	72751: contig of 24100 bp in length
72752	72851: gap of unknown length
72852	106646: contig of 33795 bp in length
106647	106746: gap of unknown length
106747	147867: contig of 41121 bp in length
147868	147967: gap of unknown length
147968	198036: contig of 50069 bp in length.
Location/Qualifiers	
1..198036	/organism="Homo sapiens"
/db_xref="taxon:9606"	
/chromosome="4"	
/clone="RP13-578N3"	
1..1156	/note="assembly_name:Contig19"
1257..2450	/note="assembly_name:Contig20"
2551..3768	/note="assembly_name:Contig24"
3869..5007	/note="assembly_name:Contig26"
5108..7155	/note="assembly_name:Contig38"
7256..8408	/note="assembly_name:Contig39"
8509..12850	/note="assembly_name:Contig42"
12951..19909	/note="assembly_name:Contig43"
20010..27734	/note="assembly_name:Contig44"
27835..48551	/note="assembly_name:Contig45"
48652..72751	/note="assembly_name:Contig46"
72852..106646	

```
/note="assembly_name:Contig47"
106747. .147867
/note="assembly_name:Contig48"
147968. .198036
/note="assembly_name:Contig49"
38208 c 37214 g 60444 t 1318 others
BASE COUNT 60792 a 38208 c 37214 g 60444 t 1318 others
ORIGIN

Query Match 7.6%; Score 59.4; DB 2; Length 198036;
Best Local Similarity 51.7%; Pred. No. 0.13;
Matches 135; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 21 taaggcctcacccctactaaaatgtgcacaaattttattatgaatagatgaagagctg 80
Db 28352 TCAGTTATTTCTATATAGCAGCTCGAGAGAAATGGCTTTTCAGATTGCATCTGAATGGC 28411

Qy 81 tagctataaattatgagagtgagtttattttatttatttcccaaatglagtccataatgc 140
Db 28412 AACATAAATAATTTAAATAATAAATTTAAATTTAAATTAATAATATGTAATTAATATAA 28471

Qy 141 ataatagcaactcactaaattcttagaataaaaaaagaataaaatgttaatttttggag 200
Db 28472 TTAATTTAATAATAATAACATGTAATTAATTTAAATTAATTTAATAATTAATAA 28531

Qy 201 gaaatggtaattttttcacaaaattgtgcagacctttacagaccttactcttcacaa 260
Db 28532 CATGTAATTAATTAATAATAATAATAATAATAACATGTAATTAATTAATAATTAATAA 28591

Qy 261 ttgacttgacattacacatca 281
Db 28592 TTAATAATAACATGTAATTAATA 28612

RESULT 7
AC006907/c
LOCUS
DEFINITION 188972 bp DNA linear HTG 26-FEB-1999
Caenorhabditis elegans clone Y76B12, *** SEQUENCING IN PROGRESS
**, 25 unordered pieces.
ACCESSION AC006907
VERSION AC006907.2 GI:4309897
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 188972)
AUTHORS Waterston,R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 188972)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (24-FEB-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Mar 1, 1999 this sequence version replaced gi:4263441.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 4051: contig of 4051 bp in length
* 4052 4061: gap of unknown length
* 4062 15145: contig of 11084 bp in length
* 15146 15155: gap of unknown length
* 15156 18371: contig of 3216 bp in length
* 18372 18381: gap of unknown length
* 18382 29860: contig of 11479 bp in length
* 29861 29870: gap of unknown length
```

```
29871 40166: contig of 10296 bp in length
* 40167: gap of unknown length
* 40177 44947: contig of 4771 bp in length
* 44948 44957: gap of unknown length
* 44958 50599: contig of 5642 bp in length
* 50600 50609: gap of unknown length
* 50610 59733: contig of 9124 bp in length
* 59734 59743: gap of unknown length
* 59744 59744: contig of 8673 bp in length
* 59745 68417: gap of unknown length
* 68418 68427: contig of 2241 bp in length
* 68428 70668: contig of 16537 bp in length
* 70669 87215: contig of 4344 bp in length
* 87216 87225: gap of unknown length
* 87226 91569: contig of 4370 bp in length
* 91570 95949: gap of unknown length
* 95950 95959: contig of 2712 bp in length
* 95960 98671: gap of unknown length
* 98672 103766: contig of 5086 bp in length
* 103767 103776: gap of unknown length
* 103777 110594: contig of 6818 bp in length
* 110595 110604: gap of unknown length
* 110605 117588: contig of 6983 bp in length
* 117589 117597: gap of unknown length
* 117598 136490: contig of 18892 bp in length
* 136491 136500: gap of unknown length
* 136501 141197: contig of 4698 bp in length
* 141198 141206: gap of unknown length
* 141207 147501: contig of 6295 bp in length
* 147502 147511: gap of unknown length
* 147512 161273: contig of 13762 bp in length
* 161274 161282: gap of unknown length
* 161283 163241: contig of 1959 bp in length
* 163242 163249: gap of unknown length
* 163250 167605: contig of 4356 bp in length
* 167606 167614: gap of unknown length
* 167615 170377: contig of 2763 bp in length
* 170378 170386: gap of unknown length
* 170387 188972: contig of 18586 bp in length.
FEATURES
Location/Qualifiers
source 1..188972
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
BASE COUNT 61066 a 32767 c 33053 g 60509 t 1577 others
ORIGIN

Query Match 7.6%; Score 59; DB 2; Length 188972;
Best Local Similarity 48.8%; Pred. No. 0.15;
Matches 219; Conservative 0; Mismatches 225; Indels 5; Gaps 2;

Qy 134 taatagcataatgacaaattcactaaatcttagaataaaaaaagaataaaatgttaatt 193
Db 50564 TAAATATATAAAAAATTTTTCACAAATTAATAATTTTTCGAAATTAATTT 50505

Qy 194 ttggagggaatgggttaatttttctcacaaaattgtgtgacagc-ttacagacattact 252
Db 50504 TTTTGAAAAATATTTTTCGAAATTAATAATTTTTCGAAATTAATTTTTCGAAAT 50445

Qy 253 cttcacaaattgacttgacacataacacagaagggctcttttcaaaaagaatgac 312
Db 50444 ATTTTTCGAAATTAATAATTTTTCGAAATTAATAATTTTTCGAAATTAATTTT 50385

Qy 313 aagaacttcgaatttttgacagtgactcttttcaacccttcaaccataatattta 372
Db 50384 TTGAAATTAATTTTTCGAAATTAATAATTTTTCGAAATTAATTTTTCGAAATTAATTT 50325

Qy 373 agtgcacatcgcttctcttcttcccaactcatttgttaactagttttctcttgagttcc 432
Db 50324 TTTTTCGAAATTAATTTTTCGAAATTAATAATTTTTCGAAATTAATAATTTTTCGAAATTAATTT 50265
```


Matches	300;	Conservative	0;	Mismatches	364;	Indels	15;	Gaps	17;
Qy	85	tataaaattgagagaagtgttattttatatttatcccaaatgtagttcataataacataa	144						
Db	12915	TAAACATATAACAATAAATAATATATATATAATAAAAAGATTCTTTCAAAATTTAT	12974						
Qy	145	tagcaacttcactaacctcagataaaaaataaaaatgttaatttttttggaggaaa	204						
Db	12975	TTTCGCATTAAACACATTCATCATTTTTAAATATATATATATATATATATATATTTT	13034						
Qy	205	tggftaatttttctacaaaaatttgtcacqcttttacagaccttactcttccacaattga	264						
Db	13035	TTTTTTTATTATTTTATTTTAAATGAGGGGTATATCTCTTAATTCACACATAGT	13094						
Qy	265	cittgaacattaacatacaaaaaggggctctgtttcaaaaaagaatagccaagaactcg	324						
Db	13095	TTTAAATAATATATTTTATTTAAATTTATTTGAIAAAAAAANAANAATGAATA	13154						
Qy	325	aatttttgacagtgactctttctaaccctttaatccaaaata-----t	369						
Db	13155	AATATATATATTTATATATATGTGATATATATATATATATATATATATATATATCCT	13214						
Qy	370	ttaagtgtccatcgctcttcctttcccaactcattgtttaaactagtttctctgtcagt	429						
Db	13215	TTACTAATTTTAAATATACACATATATAATACATATATTAATTTTGCTTTATTTTAA	13274						
Qy	430	tcctttgcctataattgacgagcttctcgaaatacccacaaactgatltcatgaagccc	489						
Db	13275	ATAAATAAACAAATAAATAATTCTGNATACAAAAATTTTAAATAATCTTAAATACTCC	13334						
Qy	490	atgcttttggaaagatttgcacttcggctcttgcaactctatttacattgactgtacttga	549						
Db	13335	TTAAATCAAGACGAAATTTAAAGAGTGATATATAACATATATATATATATATATATGTA	13394						
Qy	550	tgtattgctagatgttgactatocagttagggacaatcaaaaagatatagataatggcca	609						
Db	13395	ATTTTTTTTTTTTTTATATGAATATTTTAATATGGTATTAATAATATATTTAAATATGAAA	13454						
Qy	610	gggataaatcagaagttactgtccaatacaaaagtattgttttatgggtattttataggtg	669						
Db	13455	AATAAATCTGTTAATTTTTTGTATAGTGCAAAAATAATAAGTATGNACTTAAATAAG	13514						
Qy	670	ataaattcattactgagcaatttcataatcogtttttaattctcctcggttgtaatggtg	729						
Db	13515	AGACATTTGATTTTATTTTATTTTTTTTTTTTAAATTTTCACTATTTTCATGTGACATTTT	13574						
Qy	730	actctggagactcaaat	748						
Db	13575	TTTCTTAAAAATAAATAT	13593						

RESULT	10
PF38CBP	
LOCUS	PPF38CBP 4325 bp DNA linear INV 12-MAY-1992
DEFINITION	P.falciparum 3.8 gene for putative serine kinase (partial) and GBPj30 gene for glycophorin binding protein (partial).
ACCESSION	Z11832
VERSION	Z11832.1 GI:9837
KEYWORDS	3.8 gene; GBPj30 gene; glycophorin binding protein.
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. Lanzer M., de Bruin,D. and Ravetch,J.V. Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates
REFERENCE	EMBO J. 11 (5), 1949-1955 (1992)
AUTHORS	Lanzer,M.M.
TITLE	Direct Submission
JOURNAL	Submitted (11-MAR-1992) Lanzer M. M., Memorial Sloan-Kettering
MEDLINE	92258406
REFERENCE	2 (bases 1 to 4325) Lanzer,M.M.
AUTHORS	
TITLE	
JOURNAL	


```

CDS
join(29296..29313,29447..30361,30480..30638)
/gene="vir7"
/notes="vir7", rpf1, putative transmembrane protein, member
of rpf subfamily, similar to rpf2, vir29: MW:42155 (362
aa) fasta scores: opt: 805, E(): 1.4e-07, 39.8% identity
in 374 aa overlap, and to rpf3, vir27: MW:42372 (362 aa)
fasta scores: opt: 487, E(): 0.0004, 26.9% identity in 376
aa overlap"
/codon_start=1
/protein_id="CAB96696.1"
/db_xref="GI:8953648"
/translation="MAEILGDEKLEILPTKHNYLLDNGYTFCEKDFYNHAERKLSS
DSGLQKVFELIKAMCYMTKKRHKIFDSMDCKFFYYWLADILINLNDNHTSEVLI
NLYRIENAGAGKICDPINSYIDKNFNPCKLIFDSEYSEVYKLDLAIPNSCHENY
KKYLQYVNYKELHGYCEVEKRSDDTYCKEFKHFSNKEHLLSTWSCNLEGTGLQNL
LIDGNOKVVEEAQIGKSGTGKRGAELEKTVKRGQAEPVILTASSLETGSSLLMS
SSVEDESSISTSTSKYIANVASGAVIVFSILYINFTPAGTWINLFGRTPKMHNHLL
ADMLINNSQTSQSPNTERSFRDISIRPV"
join(34090..34134,34277..35083,35199..35366)
/gene="vir8"
join(34090..34134,34277..35083,35199..35366)
/gene="vir8"
/notes="vir8, putative transmembrane protein, low
similarity to other VVIVD10 CDS"
/codon_start=1
/protein_id="CAB96697.1"
/db_xref="GI:8953649"
/translation="MDPLSQBIDKWKDDPYPLESWKLFGEYNNDYDFBETHVELC
DFIISDEPERGNMKNYKHFICIKLLRLNHAASEITYNNMSPERCINLKWLYFYRKID
VPVKEFIRKVFNIIFNLDEIMPPYKCKYSPVEEFIESENVVKLSFLDNLSTIIDII
KFEKHLRLDCLDITRKCYEIVYTMQYCNVANKTEPSIVSLCAELAGFSESVDTH
EDFLTEGSDPVVILPLGDRSAIHRGGEKSIAESLSPLSKSTIGTAVAGCTH
GVGLGYKFTPARSWFHGRNCGVKANFLDEGEINEMFHNPNFENMESDINSYNIYG
HNMD"
join(37501..37539,37698..38378,38507..38671)
/gene="vir9"
join(37501..37539,37698..38378,38507..38671)
/gene="vir9"
/notes="vir9, rpg2, putative transmembrane protein, similar
to vir15:MW:36805 (312 aa) fasta scores: opt: 398, E():
0.00012, 28.6% identity in 289 aa overlap, and to vir12:
MW:35019 (292 aa) fasta scores: opt: 377, E(): 0.00026,
26.1% identity in 287 aa overlap"
/codon_start=1
/protein_id="CAB96698.1"
/db_xref="GI:8953650"
/translation="MAEVEKQALHYKHEIFHDSYSTNYQIFTSYCSNYKSYLQSDSE
NKCSLSESCIIICKYINLLEQNHAIISTEGCKFLYMYMDKYINSQSGCNMLPEYK
ASLSYCENGWDECNPNVEQIDVHLFEKYNLNLVLCQDFKLINSQDNEPNKCYA
KKVELVEKYIKELCGVNNYCNELKNFKHDYEKKIKELTCADIAFILTSSEFTDLA
SIIIVPVIIPALCFIYIVYVKNFTPPGWSLSLRMYRKNKYSNIENVTQYIHSRPT
EGNFQNRPNYSVNSGOY"
complement(join(40530..41207,41326..41455,41564..41634))
/gene="vir10"
complement(join(40530..41207,41326..41455,41564..41634))
/partial
/gene="vir10"
/notes="vir10, rpd2, putative transmembrane protein, start
unclear, member of rpd subfamily, similar to vir35, rpd1:
(288 aa) fasta scores: opt: 1328, E(): 0, 70.2% identity
in 292 aa overlap, and to rpd3, vir11: MW:25866 (218 aa)
fasta scores: opt: 390, E(): 8.7e-15, 34.0% identity in
206 aa overlap"

Query Match      7.1%; Score 55.6; DB 3; Length 155711;
Best Local Similarity 42.3%; Pred. No. 0.61;
Matches 304; Conservative 0; Mismatches 414; Indels 0; Gaps 0;

Oy 39 taaaatgtgatcaaaattttattatgaatagatgaaagctgtagctataaattatgaga 98
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 113247 TAAATTTTATTAATAATATATAATAATACACACATATATATAATAATATAGCAGACA 113188
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Oy 99 gtaagttttatttatcccaaatgttagttcataatagcataaatagcaaatcacta 158

```

```

Db 113187 ATAATGAAAAAATAATATAATAATAATAATAATAATAATAATAATAATAATAATA 113128
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 159 aatcttgaataaaaaaataaataaataaataaataaataaataaataaataaataaataa 218
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 113127 AATTACAATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 113068
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 219 tacaaaaattgtgacagctttacagacaccttactcttcacaaattgacttgaacattaca 278
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 113067 AACAAATATATATCTTAAATATATAGTAATAATCACTTTTAAATGTTAATAAGACAATA 113008
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 279 tcacaaagggggtccctggtttacacaaagaatagatcaagaacttcagaatttttgacagt 338
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 113007 TTGACTCATATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 112948
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 339 actcttttctaacccctttaatcccaaatatatttaagtgtccatgctcttccttccatcaa 398
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112947 TCACGATAGAAATTTATTTATGAGCATATTTTACAATATGATTAATTTTATTTATGTTT 112888
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 399 ctcattgttaactagttttcttctgtgagttcccttgcctataattgaagcagttctct 458
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112887 AATATGCTGTTTATATATATATATATATATATATATATATATATATATATATATATAT 112828
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 459 gaaatcacccaaactgatttatgaaagcccatgcttttggaaagatttgcacttcggtct 518
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112827 AATAATTTTATAGTTTCTTCTTATTTGTTGTAATGTTTAAACATAATTTTATTTTAT 112768
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 519 ttgcaatctattacattgactgactgactgactgactgactgactgactgactgactgactgact 578
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112767 AATTAAGTATTAATAATAATAAGGATCAATAATTTTAAATATACATTAATAATAATAT 112708
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 579 ggacaatcaaaaatattatgataatggcaggagataaaatcagagatttactgctcaataac 638
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112707 AGTTTATTTGATATATGTTAATAACAATAATTTTGTATAAGATGATTTTTCCTCAATA 112648
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 639 aaagtattgttttattgaggtattttattgaggtataaaattcattctgagcaattcattcatc 698
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112647 TTTCTTGAATATTTCTGTTATTTATTTTATTTTAAATATACATTAATAATAATATATAT 112588
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Oy 699 atgttttaattctcctggtgtgataatgggactcctggagactcaaatataataatt 756
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112587 TAATAATAATAATTTCTTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 112530
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 12
AC103311
LOCUS
DEFINITION
Rattus norvegicus chromosome Rfl clone DNA linear HTG 21-DEC-2001
SEQUENCE, 54 unordered pieces.
AC103311
ACCESSION
VERSION
AC103311.2 GI:17974835
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 228309)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinoh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

```


Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisghe, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogulu, M., Okwuonu, G., Oradunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaik, F., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, F., Wu, F. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 228309)
Worley, K. C.

Submitted
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:17062975.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GKEY
Center clone name: CH230-12416
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329First call to findPhrapList
Consensus quality: 206574 bases at least Q40
Consensus quality: 213453 bases at least Q30
Consensus quality: 218197 bases at least Q20
Estimated insert size: 205273; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 3.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	21216:	contig of 21216 bp in length
*	21217	21316:	gap of unknown length
*	21317	35507:	contig of 14191 bp in length
*	35508	35607:	gap of unknown length
*	45232	contig of 9625 bp in length	
*	45233	45332:	gap of unknown length
*	45333	55414:	contig of 10082 bp in length
*	55415	55514:	gap of unknown length
*	65715	contig of 10201 bp in length	
*	65716	65815:	gap of unknown length
*	65816	75695:	contig of 9880 bp in length
*	75696	75795:	gap of unknown length

*	75796	83809:	contig of 8014 bp in length
*	83810	83909:	gap of unknown length
*	83910	91085:	contig of 1776 bp in length
*	91086	91185:	gap of unknown length
*	91186	97297:	contig of 6112 bp in length
*	97298	97397:	gap of unknown length
*	97398	102349:	contig of 4952 bp in length
*	102350	102449:	gap of unknown length
*	102450	108688:	contig of 6239 bp in length
*	108689	108788:	gap of unknown length
*	108789	114011:	contig of 5223 bp in length
*	114012	114111:	gap of unknown length
*	114112	120192:	contig of 6081 bp in length
*	120193	120299:	gap of unknown length
*	120293	125370:	contig of 5078 bp in length
*	125371	125470:	gap of unknown length
*	125471	132566:	contig of 7096 bp in length
*	132567	132666:	gap of unknown length
*	132667	137559:	contig of 4893 bp in length
*	137560	137659:	gap of unknown length
*	137660	141610:	contig of 3951 bp in length
*	141611	141710:	gap of unknown length
*	141711	145979:	contig of 4269 bp in length
*	145980	146079:	gap of unknown length
*	146080	149339:	contig of 3860 bp in length
*	149940	150039:	gap of unknown length
*	150040	153903:	contig of 3866 bp in length
*	153906	154005:	gap of unknown length
*	154006	156574:	contig of 2569 bp in length
*	156575	156675:	gap of unknown length
*	160465	160564:	contig of 3790 bp in length
*	160565	164301:	contig of 3737 bp in length
*	164302	164401:	gap of unknown length
*	164402	168556:	contig of 4155 bp in length
*	168557	168657:	gap of unknown length
*	171720	171819:	contig of 3063 bp in length
*	171820	174507:	contig of 2688 bp in length
*	174508	174607:	gap of unknown length
*	174608	178199:	contig of 3592 bp in length
*	178200	178299:	gap of unknown length
*	178300	180055:	contig of 1756 bp in length
*	180056	180155:	gap of unknown length
*	180156	182612:	contig of 2457 bp in length
*	182613	182712:	gap of unknown length
*	182713	185727:	contig of 3015 bp in length
*	185728	185827:	gap of unknown length
*	185828	188739:	contig of 2912 bp in length
*	188740	188839:	gap of unknown length
*	188840	190935:	contig of 2096 bp in length
*	190936	191035:	gap of unknown length
*	191036	192644:	contig of 1609 bp in length
*	192645	192744:	gap of unknown length
*	192745	194719:	contig of 1975 bp in length
*	194720	194819:	gap of unknown length
*	194820	196523:	contig of 1704 bp in length
*	196524	196623:	gap of unknown length
*	196624	198412:	contig of 1789 bp in length
*	198413	198512:	gap of unknown length
*	198513	198847:	contig of 1335 bp in length
*	198848	199947:	gap of unknown length
*	199948	202688:	contig of 2741 bp in length
*	202689	202788:	gap of unknown length
*	202789	204565:	contig of 1777 bp in length
*	204566	204665:	gap of unknown length
*	204666	206363:	contig of 1698 bp in length
*	206364	206463:	gap of unknown length
*	206464	207903:	contig of 1440 bp in length
*	207904	208003:	gap of unknown length
*	208004	210550:	contig of 2547 bp in length
*	210551	210650:	gap of unknown length
*	210651	212113:	contig of 1463 bp in length

TITLE
JOURNAL
REFERENCE
-AUTHORS
TITLE
JOURNAL
COMMENT

```

REFERENCE
AUTHORS      3 (bases 1 to 55339)
TITLE        Waterston,R.H.
JOURNAL      Direct Submission
SUBMITTED    05-SEP-2001
JOURNAL      Genome Sequencing Center, Washington
UNIVERSITY   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
AUTHORS      4 (bases 1 to 55339)
TITLE        Waterston,R
JOURNAL      Direct Submission
SUBMITTED    09-JAN-2002
JOURNAL      Department of Genetics, Washington
UNIVERSITY   4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT      On Jan 9, 2002 this sequence version replaced gi:15624919.
-----
              Genome Center
              Center: Washington University Genome Sequencing Center
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@wustl.edu
              ----- Summary Statistics
              Center project name: H_NH00062B19

```

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseguwa, K., Woon, P.Y., Zhang, B., Frengen, E., Tateno, M., Cataneese, J.J., and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

The clone sequenced to the left is RP11-656O12; the clone sequenced to the right is RP11-328L16, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-62B19.

	FEATURES	SOURCE
1	100%	100%
2	100%	100%
3	100%	100%
4	100%	100%
5	100%	100%
6	100%	100%
7	100%	100%
8	100%	100%
9	100%	100%
10	100%	100%
11	100%	100%
12	100%	100%
13	100%	100%
14	100%	100%
15	100%	100%
16	100%	100%
17	100%	100%
18	100%	100%
19	100%	100%
20	100%	100%
21	100%	100%
22	100%	100%
23	100%	100%
24	100%	100%
25	100%	100%
26	100%	100%
27	100%	100%
28	100%	100%
29	100%	100%
30	100%	100%
31	100%	100%
32	100%	100%
33	100%	100%
34	100%	100%
35	100%	100%
36	100%	100%
37	100%	100%
38	100%	100%
39	100%	100%
40	100%	100%
41	100%	100%
42	100%	100%
43	100%	100%
44	100%	100%
45	100%	100%
46	100%	100%
47	100%	100%
48	100%	100%
49	100%	100%
50	100%	100%
51	100%	100%
52	100%	100%
53	100%	100%
54	100%	100%
55	100%	100%
56	100%	100%
57	100%	100%
58	100%	100%
59	100%	100%
60	100%	100%
61	100%	100%
62	100%	100%
63	100%	100%
64	100%	100%
65	100%	100%
66	100%	100%
67	100%	100%
68	100%	100%
69	100%	100%
70	100%	100%
71	100%	100%
72	100%	100%
73	100%	100%
74	100%	100%
75	100%	100%
76	100%	100%
77	100%	100%
78	100%	100%
79	100%	100%
80	100%	100%
81	100%	100%
82	100%	100%
83	100%	100%
84	100%	100%
85	100%	100%
86	100%	100%
87	100%	100%
88	100%	100%
89	100%	100%
90	100%	100%
91	100%	100%
92	100%	100%
93	100%	100%
94	100%	100%
95	100%	100%
96	100%	100%
97	100%	100%
98	100%	100%
99	100%	100%
100	100%	100%

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-62B19"
/clone_lib="RPC1-11"
1. .486
/note="match to EST BG483564 (NID:gl3415843)"
748. .843
/note="match to EST BG772181 (NID:gl4082834)"
748. .843
/note="match to EST BG179362 (NID:gl2686065)"
1346. .1488
repeat_region
misc_feature
misc_feature
misc_feature
repeat_region

```

SOURCE	ORGANISM
humans.	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 55339)
AUTHORS	Sulston,J.E. and Waterston,R.
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE	99063792
REFERENCE	2 (bases 1 to 55339)
AUTHORS	Bielicki,L. and Kozlowski,A.
TITLE	The sequence of Homo sapiens BAC clone RP11-62B19
JOURNAL	Unpublished (2001)

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 148983)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choepei, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,
Ferrelira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, J., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (29-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6730882.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: WBIR
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3818
Center clone name: 463_O.15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 135109 bases at least Q40
Consensus quality: 140835 bases at least Q30
Consensus quality: 143767 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 147083; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1085: contig of 1085 bp in length
* 1086 1185: gap of 100 bp
* 1186 2480: contig of 1295 bp in length
* 2481 2580: gap of 100 bp
* 2581 3989: contig of 1409 bp in length
* 3990 4089: gap of 100 bp
* 4090 4735: contig of 646 bp in length
* 4736 4835: gap of 100 bp
* 4836 6748: contig of 1913 bp in length
* 6749 6848: gap of 100 bp
* 6849 8462: contig of 1614 bp in length
* 8463 8562: gap of 100 bp
* 8563 10571: contig of 2009 bp in length
* 10572 10671: gap of 100 bp
* 10672 13999: contig of 3328 bp in length
* 14000 14099: gap of 100 bp
* 14100 17146: contig of 3047 bp in length
* 17147 17246: gap of 100 bp
* 17247 22683: contig of 5437 bp in length

* 22684 22783: gap of 100 bp
* 22784 26923: contig of 4140 bp in length
* 26924 27023: gap of 100 bp
* 27024 32598: contig of 5575 bp in length
* 32599 32698: gap of 100 bp
* 32699 40892: contig of 8194 bp in length
* 40893 40992: gap of 100 bp
* 40993 49578: contig of 8586 bp in length
* 49579 49678: gap of 100 bp
* 49679 63811: contig of 14133 bp in length
* 63812 63911: gap of 100 bp
* 63912 69287: contig of 5376 bp in length
* 69288 69387: gap of 100 bp
* 69388 82708: contig of 13321 bp in length
* 82709 82808: gap of 100 bp
* 82809 97023: contig of 14217 bp in length
* 97026 97125: gap of 100 bp
* 97126 113007: contig of 15882 bp in length
* 113008 113107: gap of 100 bp
* 113108 148983: contig of 35876 bp in length.
FEATURES
Location/Qualifiers
1..148983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-463O15"
/clone_lib="RPC1-11 Human Male BAC"
1..1085
/note="assembly_fragment"
1186..2480
/note="assembly_fragment"
2581..3989
/note="assembly_fragment"
4090..4735
/note="assembly_fragment"
clone_end:T7
vector_side:left
4836..6748
/note="assembly_fragment"
6849..8462
/note="assembly_fragment"
8563..10571
/note="assembly_fragment"
10672..13999
/note="assembly_fragment"
14100..17146
/note="assembly_fragment"
17247..22683
/note="assembly_fragment"
22784..26923
/note="assembly_fragment"
27024..32598
/note="assembly_fragment"
32699..40892
/note="assembly_fragment"
40993..49578
/note="assembly_fragment"
49679..63811
/note="assembly_fragment"
63912..69287
/note="assembly_fragment"
clone_end:SP6
vector_side:left
69388..82708
/note="assembly_fragment"
82809..97025
/note="assembly_fragment"
97126..113007
/note="assembly_fragment"
113108..148983
/note="assembly_fragment"
BASE COUNT 41111 a 30021 c 30095 g 45854 t 1902 others
ORIGIN

Query Match	7.1%	Score 55;	DB 2;	Length 148983;
Best Local Similarity	54.9%	Pred. No. 0.78;		
Matches 130;	Conservative	0;	Mismatches 105;	Indels 1;
QY	34	cctactaaaatgtagtcaaaattttattatgaataagatgaaagctgtagctataaatta	93	gene
Db	113657	CCTAAAATATGGACCTGAATTTCTTTGTAACTCATCTTCAAAGGTGAAACTAAAAATA	113598	CDS
QY	94	tgagagtaagtattattattattcatccaaatgtagttcataatagcataatgcaactt	153	
Db	113597	TGAAAATAAGCTTTATTATTATTAACACTTAAACAATGTTAGATTAAATATGCGAAAAGG	113538	
QY	154	cactaaatcttagaataaaaaaatgaataaatgttaatttttttgaggaaatggttaatt	213	
Db	113537	CACATAATCTTAAGCTGAAAGAAAAGATATG--GCTGAGTTTGTGTGAGTATTTC	113480	
QY	214	ttttctcaaaatgtgtgacagctttcacagaccttactcttcacaaatgaactgaa	270	
Db	113479	TGTTAAACTATATGCTAATACATCAACATCTTTTACTTTTACTTACTGACTCTAA	113423	
RESULT	15			
PFMAL3P7/c				
LOCUS		PFMAL3P7	253305 bp	DNA linear INV 22-MAY-2000
DEFINITION		Plasmodium falciparum MAL3P7, complete sequence.		
ACCESSION		AL008974	AL008975	AL008981
		AL010146	AL010154	AL010157
		AL010187	AL010189	AL010207
		AL010189	AL010207	AL010208
		AL139179	298556	298558
VERSION		AL034559.4	GI:8052273	
KEYWORDS		HTG; 40S Ribosomal protein S3A; acyl transferase; acylaminoacyl-peptidase; ATP-dependent RNA Helicase; elongation factor; F49C12.11-like protein; HesB-like domain protein; histone H2A variant; kinesin-related protein; N-acetylglucosamine-1-phosphate transferase; protein kinase; R-CG7 repeat; R-PA3 repeat; repl1; rep20; rifin; RNA-binding protein; stavor; T-complex protein 1 epsilon subunit; telomere; var. malaria parasite P. falciparum.		
SOURCE		Plasmodium falciparum		
ORGANISM		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE		1 (bases 1 to 253305)		
AUTHORS		Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T., Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T., Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holtroyd,S., Hornsby,T., Horrocks,P., Jajelski,K., Jessal,B., Kyes,S., McLean,J., Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A., Rajandream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S., Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and Barrell,B.G.		
TITLE		The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum		
JOURNAL		Nature 400 (6744), 532-538 (1999)		
MEDLINE		99376085		
REFERENCE		2 (bases 1 to 253305)		
AUTHORS		Oliver,K., Bowman,S., Harris,D., Lawson,D., Quail,M. and Barrell,B.		
JOURNAL		Unpublished		
REFERENCE		3 (bases 1 to 253305)		
AUTHORS		Lawson,D., Bowman,S. and Barrell,B.		
TITLE		Direct Submission		
JOURNAL		Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK		
COMMENT		On May 14, 2001 this sequence version replaced gi:2982540 gi:2982541 gi:2982544 gi:2894453 gi:2894464 gi:2894551 gi:2982556 gi:2982558 gi:2982563 gi:2982564 gi:2982566 gi:2894491 gi:2982569 gi:26595974 gi:2894502 gi:2673766 gi:2894496 gi:2982571 gi:2894588 gi:2894589 gi:2982538 gi:2982539 gi:2894380 gi:4725992. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/projects/P_falciparum. Location/Qualifiers		
FEATURES		1. 253305		
source		/organism="Plasmodium falciparum"		

```

len: 181 aa; Similarity to P. falciparum elongation factor
1-beta, (AAF27524) (276 aa), fasta scores: opt: 366, E():
1.1e-15, 53.2% identity in 141 aa overlap, revised:
shortened exon 2, Pfam: match to PF00736 EF1BD, EF-1
guanine nucleotide exchange domain Score 61.10"
/codon_start=1
/product="putative elongation factor"
/protein_id="CAB39068.2"
/db_xref="GI:8052274"
/db_xref="SPTREMBL:O97278"
/translation="MANTYDELVPLSYLYLQEGGNTSKIDQANTKKPKKKEVINKS
SLIIDIRPGYENDLDEVKLKVNITMEGLTGAKHKTPFAFLGKLOVSCVIVDDL
VNDELLETENLGLNDEQLQKKQMDDEYDEDEDEIGLVQSAAEIIISFNKL"
/8235..8240
/gene="PFC0870w, MAL3P7.3"
/notes="revised splice donor sequence for exon 2 of
PFC0870w"
8580..8585
/notes="potential splice donor sequence may indicate
further splicing of PFC0870w"
join(11431..20416,20578..20690)
/gene="PFC0875w, MAL3P7.4"
join(11431..20416,20578..20690)
/gene="PFC0875w, MAL3P7.4"
/notes="PFC0875w (MAL3P7.4), hypothetical protein, len:
3032 aa, revised: added 3' exon, possibly spliced at 5',
end"
/codon_start=1
/product="hypothetical protein, PFC0875w"
/protein_id="CAB39024.2"
/db_xref="GI:8052275"
/db_xref="SPTREMBL:O97278"
/translation="MKDFLSYAENDNLNFKLYNSEEDCLNKLVAIELKSDFKNVN
SLENDLIINHENDSNLLYLNHTNIINLKSEFNINDLYKDVTEFENFENIKLKI
NPHILKVNKTYEFNFIDIKNLKNENHYQYFNDRRKLLFFYNFVNKNVETKYS
GLLSVENVYKQKNVPSYFLGSLSPQSEGFGLGKYNALPNKYENKQYGNPFHL
YYPYVLESDDKKINKEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEKNEK
NGIQSVNNSNSNDQKKNNNNNNYYYNNPGLNITNVYKIRGVGYALLNSNEK
YFNDINLKNQFNFTNVDDLSNIFYNEMWYEFFVLEYOFNSFILYNADILK
KNLLTIPRYDELNEKREITMEYKDEKEYQVNNEYINENGVDEKGGYKTELK
LNDFTIPKPMKSMINAFDTEKKNIRFVVVFLCVLPVNCVDFINKERKINLENFL
CCLKIKYIYYSWLLFYIILFFNYIITLVVYVYVKCLMNYFILLFELFYFLMS
LFTICQFQSNSSINFIATFLFLFSSFLRIIHGASNLFFVLLIPHSFCLSL
LDFILIKNIKIDYKOLFKEFNISLHLLIGSVSVLLICILANIIYKRRKMN
MLNTRYKSKSITNDRKSKQOQTYEYGNIDYMSKYPHQLKMSNDPMLGSDSE
DTCVELDTLHKQVNSOMCLRLEKIRNDRRESNGHPYDIPMPLRNNNNNNNNNN
LNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NGSKSTLNIITIMISKDSGEINFFKNSYKKNKKDIYMDVKNRNLGCRGICNGY
IKFVWNRIFRNCNDNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NDQIKNLMDVEYKKISIFICPLVRDVIILDEPFIADIKTKLKFEPDKIKNNII
FICTHDIYEANNFANDIAVIKSGOIIFPKSKNSPKLIEYKFLNIOFNGLPNEEQTN
YLSQEKIMSAALNKKGVKSPKVANVYDGVNVRISSESSINDKKKNNNNNNDD
NNKNSVDHIDHEDILVNIENFTNRRDIIKNNIIKFIKGTRENKNCPIFFNENHI
YCTVKINELESKLKLVNKNFLTYQLTIDIIYTYIYIYIYIYIYIYIYIYIYIYI
DIKLEIDPDLFFQFFQFFQFFQFFQFFQFFQFFQFFQFFQFFQFFQFFQFFQFF
LINSTNDPDDYDINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
KFIEKFSYNTYRNNKNDENDEYDNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NFFTYYIKPTLLKLLKDLGSSSFYKFTVPLFLSGLLIICVSLFGVQVNDIDLD
YSTSSNLKQSTINFTYLVKNNNNNETLGYKNLNDNNIKNLNNNSYDKLLKLE
TKODNAEESHNNNDHNDNIKOLLENNEYNNKLIHENSFNYSNTINSLLQKYC
LKEKINYIDPLEDNIYDSINNTLEKANHENEISLGTFAFLITEKKADNSDNKEDY
ETLRTNDVYKVEEDINILNFYNTYSIHSYAYYNTSLNMLSDFONILNKKSGKNI
LDGSTYDNIKVDKVEDGNDMTLLYDKENKYNLYIKOIDKNIRKNNNNNNNNNN
VSVNNNEEDINGNINYSNNNIIYTPKNNKEENENFLVKKRIRDTLNEPNIK
FHEHMRDIFYINIVFLSVIFVCFVFFPKFNEIKRIFENPHVQYIHYFOILLLE
YLYFYIYILCLFVLVIFYDYKFLMSFFCFLLYGFNIFLSCLFSSLYLHYSILFL
FFNFIFGCIISIVYLVILSYAYNNEILNLSHLVLCIFRIFDSFALSGLVNSRLC
LNVRKRWKHIDEDIMEDMSNNNFVFCFKVYNQSLGDDNNLNSIDSGVSCDSN
SPNTSGDFVLLINCIIYLSIVYKLLQIIPSKERTDKNESEKNTNIEEIIILNNSYN
VGVRKREKKNRYTFELKYNFLSNKYEKQPKENYEKNRPFYFIQKLFNLKRGNNV
DENNTWGIYKGEKVDNSNYSIYSEIDENINNNOLNSGLKNDIYNNELNDYNDPDD
NTVDIEMQISEKSNIEKMLYKETEIDKYYKQSDDDDDNDEKNFNKISOKYILKKNLI

```

```

KMRPYKIYFSSIFNNDLNLIYFFKFFCNKNDLKGSLQTIAYGONVEINNKFNKK
NONEHISEKIQMNERDYKIELIPMSIYEHFKLILTFKNIELTNVELKYIINLLMLI
VNLKCDIHLCNOLSGMKKKVELMLNLRDDKIIIFLYKLNNDIDFCSOIYINILKN
ILLYNCEDEHDEYKDYKDYKDYKDYKDYKDYKDYKDYKDYKDYKDYKDYKDYKDY
RYVEQDIQYKGGIYNDENQDFDYIDKYSYLDLGRKEDRQLRENMLKGRHGKNNY

```

```

Query Match      7.1%; Score 55; DB 3; Length 253305;
Best Local Similarity 46.1%; Pred. No. 0.69;
Matches 184; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 37 actaaatgtagcaaaattttattgataagatgaagaagctgtagctataaattatga 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198774 AATTACAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 198715

QY 97 gactaagtattttattttattttccaaatgtagttcataatagcataatagaacttcac 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198714 TATATATATATATATATATATATATATATATATATATATATATATATATAT 198655

QY 157 taaatcattagaataaaatgataaaatgtaattttttggaggaatggtaattttt 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198654 TGTATATTTTAAATATATCTATCTATGAATATTTTATAAATTTATAGTAAACGAAAAATAAT 198595

QY 217 tctacaaaattgtgtagacagctttacacagcttactcttcacaaatgactggaacattaa 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198594 AAGAATAACGCATGCACATATATATAGAAATATATCTATGTATATATAGTTGTATATAATAG 198535

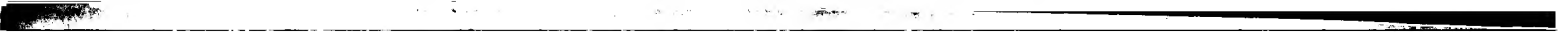
QY 277 catcacaaagagggtctgtgtttacaaaagaatagtcagaacacctgaatttttgacag 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198534 GATAATAAATAACATATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 198475

QY 337 tgactcttttctaaccttcaatccaaatataatttaagtgccatcgctcccttattcc 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198474 TAATTATTTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 198415

QY 397 aactcattgttaactagtttcttctctgtgagttccttt 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198414 TTTTAAATAATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTT 198376

```

Search completed: May 13, 2002, 22:57:25
Job time: 5020 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 21:35:30 ; Search time 331.61 Seconds
(without alignments)
4038.459 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgttctttaaag.....taaaagcaaaaaa 780

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

```

1: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780	100.0	780	22	AAD19683
2	50.8	6.5	5278	22	AA546375
3	50.8	6.5	5278	24	ABL32822
4	50.2	6.4	11422	24	ABL32218
5	49.8	6.4	4590	7	AA60472
6	49.2	6.3	6179	22	AA546343
7	48.6	6.2	5516	23	ABL19458
8	48.4	6.2	17491	24	ABL34575
9	48.2	6.2	6113	24	ABL32430

10	48.2	6.2	19087	24	ABL32793	Human immune syste
11	47.8	6.1	6191	24	ABL33217	Human immune syste
12	47.8	6.1	18218	24	ABL33949	Human immune syste
13	47.4	6.1	5947	22	AA546675	Tumour suppressor
14	47.4	6.1	6106	22	AA546430	Tumour suppressor
15	47.4	6.1	6106	24	ABL33473	Human immune syste
16	47.4	6.1	16918	24	ABL33617	Human immune syste
17	47.2	6.1	4163	24	ABL34330	Human immune syste
18	47	6.0	5237	24	ABL34617	Human metastasis a
19	47	6.0	12356	22	AA546509	Tumour suppressor
20	47	6.0	13326	24	ABL33713	Human immune syste
21	46.8	6.0	13377	22	AA546476	Tumour suppressor
22	46.8	6.0	13377	24	ABL33463	Human immune syste
23	46.4	5.9	912	21	AA53727	Campylobacter jeju
24	46.4	5.9	7329	22	AA546673	Tumour suppressor
25	46.4	5.9	21537	24	ABL33398	Human immune syste
26	46.2	5.9	53585	20	AA20251	Borrelia burgdorfe
27	46	5.9	6988	24	ABL34441	Human immune syste
28	46	5.9	15587	24	AA561238	Human gene regulat
29	45.8	5.9	500	21	AA593655	Cat flea head and
30	45.8	5.9	6106	22	AA546429	Tumour suppressor
31	45.8	5.9	6106	24	ABL33472	Human immune syste
32	45.8	5.9	11812	22	AA545502	Chemically pretrea
33	45.8	5.9	11812	24	AA546742	Tumour suppressor
34	45.8	5.9	11812	24	ABL34119	Human immune syste
35	45.8	5.9	15872	22	AA546519	Tumour suppressor
36	45.8	5.9	18011	24	ABL32034	Human immune syste
37	45.6	5.8	5841	24	ABL33396	Human immune syste
38	45.6	5.8	5860	24	AA561372	Human gene regulat
39	45.6	5.8	6852	24	AA561249	Human gene regulat
40	45.6	5.8	15373	24	ABL32467	Human immune syste
41	45.4	5.8	6052	24	AA561389	Human gene regulat
42	45.4	5.8	6118	24	ABL33031	Human immune syste
43	45.4	5.8	8588	22	AA545470	Chemically pretrea
44	45.4	5.8	16602	24	ABL32726	Human immune syste
45	45.2	5.8	6076	22	AA546664	Tumour suppressor

ALIGNMENTS

RESULT 1

AAD19683

ID AAD19683 standard; cDNA; 780 BP.

XX AAD19683;

XX 18-DEC-2001 (first entry)

XX Human Mammary Gland Cancer Specific Gene (MSG) mam021 cDNA.

XX Human; Mammary Gland Cancer Specific Gene; MSG; cytostatic; vaccine;
cancer; therapy; immune response; ss.

OS Homo sapiens.

XX WO200172780-A2.

PN 04-OCT-2001.

PD 26-MAR-2001; 2001WO-US09525.

PF 27-MAR-2000; 2000US-192277P.

XX (DIAD-) DIADEXUS INC.

XX Salceda S, Hu P, Recipon H, Caferkey R;

XX WPI; 2001-616468/71.

XX New isolated polynucleotide, mammary gland cancer specific gene (MSG),
useful for diagnosing, monitoring, staging, imaging and treating
mammary gland cancer

XX

PS Claim 1; Page 87; 99pp; English.

CC The present sequence is human mammary gland cancer specific gene (MSG)
 CC cDNA. MSG is useful for diagnosing, detecting, monitoring, staging,
 CC prognosticating, imaging and treating mammary gland cancer in a patient
 CC by determining the levels of MSG in cells, tissues or bodily fluids in a
 CC patient and comparing the determined levels of MSG with levels of MSG
 CC in cells, tissues or bodily fluids from a normal human control, where a
 CC change in determined levels of MSG in the patient versus normal control
 CC is associated with the presence of mammary gland cancer. MSG is used for
 CC identifying potential therapeutic agents for use in imaging and treating
 CC mammary gland cancer. MSG antibody conjugated to a cytotoxic agent is
 CC useful for treating mammary gland cancer in a patient. MSG vaccine is
 CC useful for inducing an immune response against a MSG protein and for
 CC treating mammary gland cancer in a patient. MSG and its protein are
 CC useful as diagnostic markers for mammary gland cancer and for diagnosis
 CC and treatment of disorders of cells, tissues and organisms.

XX Sequence 780 BP; 267 A; 115 C; 119 G; 279 T; 0 other;

Query Match 100.0%; Score 780; DB 22; Length 780;
 Best Local Similarity 100.0%; Pred. No. 1.2e-146;
 Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aattactgttctcttaaaagtaagccttacacccctactataaattgtgatcaaaattttat 60

Db 1 aattactgttctcttaaaagtaagccttacacccctactataaattgtgatcaaaattttat 60

QY 61 tatgaataagatgaagcgtgtagctataaaattatgagagtaagttattttattatc 120

Db 61 tatgaataagatgaagcgtgtagctataaaattatgagagtaagttattttattatc 120

QY 121 caaattagttcctaataagcataagcaactcactaaattcctagaataaaatgaat 180

Db 121 caaattagttcctaataagcataagcaactcactaaattcctagaataaaatgaat 180

QY 181 aaaaatttaatttttggaggaaatggttaattttttacaaaattgtgcagcgttt 240

Db 181 aaaaatttaatttttggaggaaatggttaattttttacaaaattgtgcagcgttt 240

QY 241 acagaccctactctcacaattgacttgaacatttaacatacaagaagggctcgtttac 300

Db 241 acagaccctactctcacaattgacttgaacatttaacatacaagaagggctcgtttac 300

QY 301 aaaaagatagtcagaacttcgatatttttgacagtgactcttttcaaccccttaate 360

Db 301 aaaaagatagtcagaacttcgatatttttgacagtgactcttttcaaccccttaate 360

QY 361 caaatattttaagtgctcatgcttcttcttaccactcatttggtaactagttttct 420

Db 361 caaatattttaagtgctcatgcttcttcttaccactcatttggtaactagttttct 420

QY 421 tctgtagttctcttgcctataaattgaagcaggttctctgaaatcaccacaactgaattta 480

Db 421 tctgtagttctcttgcctataaattgaagcaggttctctgaaatcaccacaactgaattta 480

QY 481 tgaagcccatgcttttggaaagatttgcaacttggcttggcttgcactctatttaccattgact 540

Db 481 tgaagcccatgcttttggaaagatttgcaacttggcttggcttgcactctatttaccattgact 540

QY 541 gtacttgcattgtatgttagatttcaactatcagtttaggacaatacaaaagatattaga 600

Db 541 gtacttgcattgtatgttagatttcaactatcagtttaggacaatacaaaagatattaga 600

QY 601 taatggcgagggtataatcagaagttactgtcaatacaaaagttattttatgggtatt 660

Db 601 taatggcgagggtataatcagaagttactgtcaatacaaaagttattttatgggtatt 660

QY 661 ttatagtgataaattcattactagcgaatttcaatacatcatttttaattctctgtgtgt 720

Db 661 ttatagtgataaattcattactagcgaatttcaatacatcatttttaattctctgtgtgt 720

QY 721 aatatggtgactctggagactcaaatatttaattggtgtaaaaggcaaaaaa 780
 Db 721 aatatggtgactctggagactcaaatatttaattggtgtaaaaggcaaaaaa 780

RESULT 2

AA546375
 ID AAS46375 standard; DNA; 5278 BP.

XX AAS46375;

XX 18-DEC-2001 (first entry)

XX Tumour suppressor gene derived chemically modified sequence #97.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.

XX Homo sapiens.

XX WO200168912-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2001-602752/68.

XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 cancer

XX Claim 1; SEQ ID No 97; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5278 BP; 1568 A; 84 C; 1035 G; 2591 T; 0 other;

KW Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
XX
PR 06-APR-2000; 2000DE-1019058.
XX
PR 07-APR-2000; 2000DE-1019173.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumour suppressor
XX genes and oncogenes, useful in designing primers and probes for
XX PT analysing diseases associated with cytosine methylation state e.g.
XX cancer -
XX
XX Claim 1; SEQ ID No 65; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and
XX oncogenes having a sequence taken from 536 (actually 533 since
XX numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX (ss) and sequences complementary to (Ss). The nucleic acid may be a
XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX form part of a set of probes for detecting the cytosine methylation state
XX and/or single nucleotide polymorphisms and also to be used in an
XX array for analysing diseases associated with CpG dinucleotides e.g.
XX cancers and tumours. The probes can also be used in a method for
XX ascertaining genetic and/or epigenetic parameters for the diagnosis
XX and/or therapy of existing diseases or the predisposition to specific
XX diseases, by analysing cytosine methylations. The parameters may be
XX compared to another set of genetic and/or epigenetic parameters, the
XX differences serving as basis for diagnosis and/or prognosis events which
XX are disadvantageous to patients. The present sequence is one of the
XX 533 genomic sequences derived from tumour suppressor genes and
XX oncogenes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6179 BP; 1761 A; 105 C; 1302 G; 3011 T; 0 other;
SQ

Query Match 6.3%; Score 49.2; DB 22; Length 6179;
Best Local Similarity 51.4%; Pred. No. 0.59;
Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
0;

QY 5 actgtgtctcttaagtaaggccctacacccctactaaaatgtagatcaaaattttattatg 64
DB 2832 AATTATTCCTTAAATAAATACATTCTCTAAAATACACATTATATAAAAATTTCTTTAA 2773
QY 65 aatagatgaagaagctgtagctataaaattatgagagtaagttttattttattatccaaa 124
DB 2772 AACTATTTTAACTTCCAAATTTCAATTTTCAAAATAAATCAATTCATTCCTTTATTTT 2713
QY 125 ttagtgcataatagcatatagcaactcactcaaatcttagaataaaaaatgaataaaa 184
DB 2712 TTAATCTATATTTAAAAAATAACAAATATTTAAAAAATAAATAAATAAATAAATAA 2653

QY 185 tgttaatttttggagaaatggtaattttttttctacaaaat 226
DB 2652 AATAAACTTTAAAAAACAATCCCAATCTATCTTCAAACT 2611

RESULT 7
ABLL9458/C
ID ABL19458 standard; DNA; 5516 BP.
XX
XX ABL19458;
AC
XX
XX 26-MAR-2002 (first entry)
DT
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 9847.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 9847; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 5516 BP; 1766 A; 885 C; 914 G; 1951 T; 0 other;
SQ

Query Match 6.2%; Score 48.6; DB 23; Length 5516;
Best Local Similarity 46.0%; Pred. No. 0.77;
Matches 165; Conservative 0; Mismatches 194; Indels 0; Gaps 0;
0;

QY 19 agtaagcgccctacacccctactaaaatgtagatcaaaattttattatgaatagatgaaagc 78
DB 365 AGTAAATTCCTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 306
QY 79 ttagctataaaattatgagagtaagttttattttattttatccaaatgtagttcataata 138
DB 305 TATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 246
QY 139 gcataatagcaacttcactaaatcttagaataaaaaatgaataaaatgttaattttgg 198
DB 245 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 186

```

QY 199 aggaatggttaattttttctacaaaattgtgtgacagcttttacagaccttactcttcac 258
    || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 TATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 126
    || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 259 aattgactgaacattacacacacaaaggggtcctgtttacaaaagaatagtcagaac 318
    || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 125 ATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 66
    || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 319 ttcatgaatttttgacagtgactcttttctaacacctttaaaccataatatttaagtgt 377
    || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 TATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7
    || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
ABL34575/c
ID ABL34575 standard; DNA; 17491 BP.
XX AC
AC ABL34575;
XX DT
DT 26-MAR-2002 (first entry)
XX DE
DE Human metastasis associated gene SEQ ID NO: 128.
XX KW
KW Metastasis associated gene; cytostatic; gene therapy; cancer;
KW cytosine methylation; gene; ds.
XX OS
OS Homo sapiens.
XX PN
PN WO200177376-A2.
XX PD
PD 18-OCT-2001.
XX PF
PF 06-APR-2001; 2001WO-EP03970.
XX PR
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX PA
PA (EPIG-) EPIGENOMICS AG.
XX PI
PI Olek A, Piepenbrock C, Berlin K;
XX PT
PT WPI; 2002-010922/01.
XX PS
PS New nucleic acid derived from chemically treated metastasis genes,
    useful for diagnosis of cancers by analysis of cytosine methylation,
    also for treatment.
XX CC
CC Claim 1; SEQ ID NO 128; 23pp + Sequence Listing; English.
XX CC
CC The present invention provides a number of human metastasis associated
    genes which are modified by cytosine methylation. The sequences can
    be used in the diagnosis and treatment of cancer. The present sequence is
    one of the genes of the invention.
XX SQ
SQ Sequence 17491 BP; 4676 A; 284 C; 4226 G; 8305 T; 0 other;

Query Match 6.2%; Score 48.4; DB 24; Length 17491;
Best Local Similarity 51.9%; Pred. No. 0.96;
Matches 109; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 17 aaagttaagccctacacccactactaaaatgtgatcaaaattttattatgaatagatgaaa 76
    ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17291 AAAAAAATATTTCTTAAATCCAAAAAATTAATAATTTTATAAAAAACATCCCATTA 17232
    ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 gctgtagctataaattatgagagtaagttattttattttattttatccaaatgtagttcataa 136
    ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17231 ACTCTCTAAACGACTATTTCAATTTTATTTATTTTATTTTACTAATAATACTCTATAA 17172
    ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 137 tagcctaataagcaacttcactaaatccttagaataaaaaatgaataaaatgttaattttt 196
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
    , - ,

```

```

Db 17171 ATAATTATTTTAAACTAAATACTCCTTAACATAATTTTAAATAATAACTTAACATTTT 17112
QY 197 ggaggaaatggttaattttttctacaaaat 226
    ||| ||| ||| || | | | | | | | | | | | | | | | | | | | | | |
Db 17111 TCATCAAAAATTTAAATAAATTTACAAAAAATT 17082
    ||| ||| ||| || | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
ABL32430
ID ABL32430 standard; DNA; 6113 BP.
XX AC
AC ABL32430;
XX DT
DT 26-MAR-2002 (first entry)
XX DE
DE Human immune system associated gene SEQ ID NO: 403.
XX KW
KW Human; immune system disease; cytosine methylation; antiasthmatic;
    antiarteriosclerotic; antianaemic; cytostatic; nootropic;
    neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
    antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
    antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
    acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
    neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
    gene; ds.
XX OS
OS Homo sapiens.
XX PN
PN WO200200928-A2.
XX PD
PD 03-JAN-2002.
XX PF
PF 02-JUL-2001; 2001WO-EP07537.
XX PR
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX PA
PA (EPIG-) EPIGENOMICS AG.
XX PI
PI Olek A, Piepenbrock C, Berlin K;
XX PT
PT WPI; 2002-130909/17.
XX PS
PS Nucleic acid comprising fragment of chemically modified gene, useful
    for diagnosis and treatment of diseases associated with abnormal
    cytosine methylation.
XX CC
CC Claim 1; SEQ ID NO 403; 32pp + Sequence Listing; German.
XX CC
CC The present invention provides a number of human immune system associated
    genes which are modified by the methylation of cytosines. The sequences
    can be used in the diagnosis and treatment of immune system disorders,
    including eye diseases such as retinopathy, neovascular glaucoma and
    macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
    leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
    rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
    diseases. The present sequence is a gene of the invention.
XX SQ
SQ Sequence 6113 BP; 1998 A; 126 C; 1299 G; 2690 T; 0 other;

Query Match 6.2%; Score 48.2; DB 24; Length 6113;
Best Local Similarity 48.6%; Pred. No. 0.94;
Matches 161; Conservative 0; Mismatches 168; Indels 2; Gaps 1;

QY 3 ttactgtctcttaaggaagccctacacccactactaaaatgtgatcaaaattttatta 62
    ||| ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3316 ttatttattgttaaaaatacaataaattgtattattataaagggtgaatgtattgtatgt 3375
    ||| ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 tgaatagatgaagactgactataaaattatgagagtaagttattttatttatcccca 122
    ||| ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3376 aaattatatttaaaaaaatttagtttagtttagtttagtttagtttagtttagtttag 3435
    ||| ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |

```


CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pt_sequences.
 XX
 SQ Sequence 6106 BP; 2270 A; 30 C; 1049 G; 2757 T; 0 other;

Query Match 6.1%; Score 47.4; DB 22; Length 6106;
 Best Local Similarity 56.5%; Pred. No. 1.4;
 Matches 108; Conservative 0; Mismatches 81; Indels 2; Gaps 1;
 Qy 36 tactaaaaatgtgatcaaaattttattatgaatagatgaaaagctgtagctataaaattatg 95
 Db 3205 tataaattattataataattataataataataataataa--tatatttaatttaatt 3262
 Qy 96 agagtaagtgtttattttattccaaatgtagttcacaatagacataaagcaacttca 155
 Db 3263 ataataatttaataataattataataataataataataataattattattattat 3322
 Qy 156 cttaaatcttagaataaaaaatgaaataaattgttaatttttgggggaaatgggttaatttt 215
 Db 3323 attaatataataataataattataattataataataataataataataataatt 3382
 Qy 216 ttctacaaaaat 226
 Db 3383 taatataaat 3393

RESULT 15
 ID ABL33473 standard; DNA; 6106 BP.

XX ABL33473;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1446.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosine; cytosine; cytosine;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiarthritis; antidiabetic; antidiabetic; antipsoriasis;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX

PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 1446; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 6106 BP; 2270 A; 30 C; 1049 G; 2757 T; 0 other;

Query Match 6.1%; Score 47.4; DB 24; Length 6106;
 Best Local Similarity 56.5%; Pred. No. 1.4;
 Matches 108; Conservative 0; Mismatches 81; Indels 2; Gaps 1;
 Qy 36 tactaaaaatgtgatcaaaattttattatgaatagatgaaaagctgtagctataaaattatg 95
 Db 3205 tataaattattataataattataataataataataataa--tatatttaatttaatt 3262
 Qy 96 agagtaagtgtttattttattccaaatgtagttcacaatagacataaagcaacttca 155
 Db 3263 ataataatttaataataattataataataataataataataattattattattat 3322
 Qy 156 cttaaatcttagaataaaaaatgaaataaattgttaatttttgggggaaatgggttaatttt 215
 Db 3323 attaatataataataataattataattataataataataataataataataatt 3382
 Qy 216 ttctacaaaaat 226
 Db 3383 taatataaat 3393

Search completed: May 13, 2002, 22:53:43
 Job time: 4693 sec



Db 2063 TTAATTTGACATTTTAAATCGAGTTAATGTTGTTATTTTCGTTATACCAATTCATAA 2122
 Qy 159 aatcttagaataaaatgaataaagttaattttttggagggaatggtaatttttc 218
 Db 2123 AATATTTTATATTAATAATATAGCATCTCAGATGTGGTGAAGTAAATATTTA 2182
 Qy 219 taaaaatgtgtacagcctttacagaccccttaactcttcacaaattgacctgaacattaa 278
 Db 2183 ACAAATATATTTGAAAAATGATAAAATACTAAATGAGGTTTTGGTTGAATAGTAAGA 2242
 Qy 279 t 279
 Db 2243 t 2243

RESULT 2
 US-08-998-416-186
 ; Sequence 186, Application US/08998416
 ; Patent No. 6239264
 ; GENERAL INFORMATION:
 ; APPLICANT: Philippsen, Peter
 ; APPLICANT: Pohlmann, Rainer
 ; APPLICANT: Steiner, Sabine
 ; APPLICANT: Mohr, Christine
 ; APPLICANT: Wendland, Jurgin
 ; APPLICANT: Knechtle, Philipp
 ; APPLICANT: Rebischung, Corinne
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1152
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6239264artis Corporation
 ; STREET: 3054 Cornwallis Road
 ; CITY: Research Triangle Park
 ; STATE: No. 6239264th Carolina
 ; COUNTRY: USA
 ; ZIP: 27709

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 186:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 615 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1074RP
 US-08-998-416-186

Query Match 5.6%; Score 43.8; DB 4; Length 615;
 Best Local Similarity 44.5%; Pseq No. 0.18;
 Matches 215; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

QY 1 aattaattgtctcttaagtaaggccttacacccctactaaatgtgatacaaaattttat 60
 Db 58 ATTTAAGTATTAATTAATAACTATTATTATTCATTAATTAATAATTAATTAATTTGAT 117
 QY 61 tatgaatagatgaagaagcgtgtagctataaattatgatgagagtaagtktattttatttc 120
 Db 118 TATTAATACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 177
 QY 121 caaatgtagtctcataatagcataatagcaacttcacataactcttagaataaaaaatgaat 180
 Db 178 TATAAATATAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 237
 QY 181 aaaaatgttaatttttggagggaatggtaattttttcacaaaaattgtgtgacagcttt 240
 Db 238 TAGTTTATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 297
 QY 241 acagaccccttactcttcacaaattgacctgaacatttaacataacacacagagggctgtttac 300
 Db 298 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATAA 357
 QY 301 aagaataatagtcagaacattcatgaatttttgacagtgactcttttctaaccccttaacc 360
 Db 358 GATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATAA 416
 QY 361 caaatatatttaagtgtccatcgctcttctcttattccaaactcatttggtaactagttttct 420
 Db 417 ATATTATGTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 476
 QY 421 tctgtgagtcctcttcctataatgaagcagtgctctctgaaatcaccacacaaactgatttta 480
 Db 477 ACTTTAAATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 536
 QY 481 tga 483
 Db 537 TTA 539

RESULT 3
 US-08-299-953-1/c
 ; Sequence 1, Application US/08299953
 ; Patent No. 5646333
 ; GENERAL INFORMATION:
 ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
 ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
 ; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/299,953
 ; FILING DATE: Herewith
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Beardell, Lori Y.
 ; REGISTRATION NUMBER: 34,293
 ; REFERENCE/DOCKET NUMBER: NOVA-0003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-564-8960
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2861 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double

; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 US-08-299-953-1

Query Match 5.5%; Score 42.8; DB 1; Length 2861;
 Best Local Similarity 54.4%; Pred. No. 0.38;
 Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 agatgaaagcttagctataaaattatggagtgtaagtgttttatttatcccaaatgt 127
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1232 AAAAAATATCTTTATATATAAATGGAAGATAAAATTATATATAATTAACCAAAA 1173

QY 128 agtcataatagcataatagcaacttcacataaaccttagaataaaaaatgaataaaatgt 187
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1172 AGATAATAATAAAATCTAGAGTGTATAATAAAACACTAATATTAATCTTTTTAACATTG 1113

QY 188 taatttttggaggaaatgggttaattttttctcaca aaa 225
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1112 TAAATGATTATATTATGATATAATTTTTTTTCAAAA 1075

RESULT 4
 US-08-459-415-1/c
 ; Sequence 1, Application US/08459415
 ; Patent No. 5744334
 ; GENERAL INFORMATION:
 ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
 ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
 ; EXPRESSION OF FOREIGN PROTEINS TO THE PLANT EPIDERMIS
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION NUMBER: US/08/459,415
 ; FILING DATE: 02-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/299,953
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Beardell, Lori Y.
 ; REGISTRATION NUMBER: 34,293
 ; REFERENCE/DOCKET NUMBER: NOVA-0003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-564-8960
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2861 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 US-08-459-415-1

Query Match 5.5%; Score 42.8; DB 1; Length 2861;
 Best Local Similarity 54.4%; Pred. No. 0.38;
 Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 agatgaaagcttagctataaaattatggagtgtaagtgttttatttatcccaaatgt 127
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1232 AAAAAATATCTTTATATATAAATGGAAGATAAAATTATATATAATTAACCAAAA 1173

QY 128 agtcataatagcataatagcaacttcacataaaccttagaataaaaaatgaataaaatgt 187
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 1172 AGATAATAATAAAATCTAGAGTGTATAATAAAACACTAATATTAATCTTTTTAACATTG 1113

QY 188 taatttttggaggaaatgggttaattttttctcaca aaa 225
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1112 TAAATGATTATATTATGATATAATTTTTTTTCAAAA 1075

RESULT 5
 US-09-066-687-1/c
 ; Sequence 1, Application US/09066687
 ; Patent No. 6339185
 ; GENERAL INFORMATION:
 ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
 ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
 ; EXPRESSION OF FOREIGN PROTEINS TO THE PLANT EPIDERMIS
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185ris
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/066,687
 ; FILING DATE: Herewith
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Beardell, Lori Y.
 ; REGISTRATION NUMBER: 34,293
 ; REFERENCE/DOCKET NUMBER: NOVA-0003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-564-8960
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2861 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 US-09-066-687-1

```

RESULT 7
US-08-299-953-2/c
; Sequence 2, Application US/08299953
; Patent No. 5646333
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

```

```

RESULT      8
US-08-459-415-2/c
; Sequence 2, Application US/08459415
; Patent No. 5744334
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,415
; FILING DATE: 02-JUN-1995

```



```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-459-415-2

Query Match          5.5%; Score 42.8; DB 1; Length 3881;
Best Local Similarity 54.4%; Pred. No. 0.39;
Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 agatgaaagctgtagctataaattatgagagtaagtttatatttatattatccaaatgt 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1232 AAAAAAATCTTTATATATAATAAATGGAGAGATAAATTTATATATTAATAAAAAA 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 agttcataatagcataatagcaacttcactaaatcttagaataaaaaatgaataaatgt 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1172 AGATAATAATAAATCTAGAGTTATATAATAAAAAACTAATATTAATCTTTTAACATTG 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 188 taatttttggaggaaatgggttaatttttctacaaaa 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 TAAAAATGATTTATATATATGATATAATTTTTTTTCAAAA 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-066-687-2/c
; Sequence 2, Application US/09066687
; Patent No. 6339185
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,687
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-11231-2
```

```
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-066-687-2

Query Match          5.5%; Score 42.8; DB 4; Length 3881;
Best Local Similarity 54.4%; Pred. No. 0.39;
Matches 86; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 agatgaaagctgtagctataaattatgagagtaagtttatatttatattatccaaatgt 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1232 AAAAAAATCTTTATATATAATAAATGGAGAGATAAATTTATATATTAATAAAAAA 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 agttcataatagcataatagcaacttcactaaatcttagaataaaaaatgaataaatgt 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1172 AGATAATAATAAATCTAGAGTTATATAATAAAAAACTAATATTAATCTTTTAACATTG 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 188 taatttttggaggaaatgggttaatttttctacaaaa 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1112 TAAAAATGATTTATATATATGATATAATTTTTTTTCAAAA 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
PCT-US95-11231-2/c
; Sequence 2, Application PC/TUS9511231
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
; TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11231
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-11231-2

Query Match          5.5%; Score 42.8; DB 5; Length 3881;
```

	Query Match	Score 5.3%	DB 41.4;	Mismatches 167;	Indels 0;	Gaps 0;
	Best Local Similarity	45.5%;	Pred. No. 0.61;			
	Matches 147;	Conservative				
QY	11	tctcttaaaagtaaaggcccttacaacctactaaaatgtgatcaaaaaatttttatgataaga	70			
Dd	172	TTTTATATTAAATAATTATTAATAATTTTTATAAATAATTTATTAATTAATAATA	231			
QY	71	tgaagaagcgttagctataaaaattatggagtgaaagtcttttatatttcaccaaatgtagt	130			

```

RESULT 12
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573d Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29.009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; IS-08-446-855A-1

```

Query Match	5.3%;	Score 41.2;	DB 2;	Length 8920;
Best Local Similarity	53.0%;	Prod. No. 0.99;		
Matches 88;	Conservative	0;	Mismatches 78;	Indels 0;
Gaps 0;				
<i>y</i>	59	attatgaatgagtagaaaagctgtagctataaattatgagtgagtgaaagttatttttatattta	118	
<i>b</i>	8481	ACTATTAATATATAAATAGTAAATATATATATATATATATATATATATATATATATA	8540	
<i>y</i>	119	tccaaatgtagtctcatatagcatatagcagaacttcactaaatcttagaataaaaaatga	178	

```

COMPUTER: IBM COMPACT
OPERATING SYSTEM: DOS

```

Dh 459 ATATATAATATATATATATTCGTCCTTAAACCA 501

```

COMPUTER: IBM COMPACT
OPERATING SYSTEM: DOS

```

RESULT 15
US-08-605-106-4
; Sequence 4, Application US/08605106
; Patent No. 5910631
; GENERAL INFORMATION:
; APPLICANT: Topfer, R.
; APPLICANT: Martini, N.
; APPLICANT: Schell, J.
; TITLE OF INVENTION: MEDIUM CHAIN-SPECIFIC THIOESTERS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwedman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605.106
; FILING DATE: 23-SEPT-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/02935
; FILING DATE: 01-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 235.001US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-6900
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4098 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: : DNS (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Cuphea lanceolata
; IMMEDIATE SOURCE:
; LIBRARY: genomic Lambda FIX II
; CLONE: ClTeg1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1797..2294, 2658, 2791, 2898..3011, 3132
; LOCATION: ..3303, 3391..3459, 3672..3941)
; FEATURE:
; NAME/KEY: Startcodon
; LOCATION: 1797..1799
; FEATURE:
; NAME/KEY: exon II
; LOCATION: 1787..2294
; FEATURE:
; NAME/KEY: intron II
; LOCATION: 2295..2657
; FEATURE:
; NAME/KEY: exon III
; LOCATION: 2658..2791
; FEATURE:
; NAME/KEY: intron III
; LOCATION: 2792..2897
; FEATURE:
; NAME/KEY: exon IV
; LOCATION: 2898..3011

; FEATURE:
; NAME/KEY: intron IV
; LOCATION: 3012..3131
; FEATURE:
; NAME/KEY: exon V
; LOCATION: 3132..3303
; FEATURE:
; NAME/KEY: intron V
; LOCATION: 3304..3390
; FEATURE:
; NAME/KEY: exon VI
; LOCATION: 3391..3459
; FEATURE:
; NAME/KEY: intron VI
; LOCATION: 3460..3671
; FEATURE:
; NAME/KEY: exon VII
; LOCATION: 3672..3941
; FEATURE:
; NAME/KEY: Stopcodon
; LOCATION: 3942..3944
; US-08-605-106-4

Query Match 5.2%; Score 40.6; DB 2; Length 4098;
Best Local Similarity 44.9%; Pred. NO. 1.2;
Matches 154; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY 56 ttattatgaatgagaaagctgtagtacataataaattatgagagtaagttttttatat 115
Db 159 TATATTGTGAATATTTTATGAATTTATTTGAAATATTTTGGAAATTTTAAATATTTT 218
QY 116 ttatccaaatgtagttcataatagcataatagcaacttcactaaatccttagaataaaaaa 175
Db 219 AAATTTAAATATATATTTTAAATTTCTTTTAAAAAAATATTTTAAATATTTTAAATAGT 278
QY 176 tgaataaaatgttaatttttggaggaatggttaatttttttcaaaaattgtgtgaca 235
Db 279 TTTTAAATTTTAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTGAAT 338
QY 236 gcattacagaccttactcttcacaaattgaccttgacatttaacatacacaagagggtcctg 295
Db 339 TTTTAAATATATTTTGTGTTTAAAAATATATTTTAAAGTTTAAATATTTTAAATTTTGAAT 398
QY 296 ttacaaaagaatagtcacagaacttcacgaatttttgacagtgactcttttcaaccctt 355
Db 399 TTGCAATATTTGAAAAATTTTGTGGAGATAACCGGAGAAATTTATATATATATATATAT 458
QY 356 taatccaaatatatttaagtgtccatcgtcttcttctatccaa 398
Db 459 ATATATATATATATATATTTTGTCCATTTCGGTTAAACCAA 501

Search completed: May 13, 2002, 22:47:34
Job time: 4364 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 21:32:40 ; Search time 1948.16 Seconds

(without alignments)

5403.881 Million cell updates/sec

Title: US-09-817-318-1

Perfect score: 780

Sequence: 1 aattactgttcttcttaag.....taaaggcaaaaaaaaaaaaaa 780

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67.2	8.6	1101	12	CNS0039G
2	62.8	8.1	1101	12	CNS00507
3	61.4	7.9	524	12	CNS01090
4	61.2	7.8	928	12	CNS000KY
c	60.8	7.8	1101	12	CNS0042W
6	60.2	7.7	1101	12	CNS00EVL
c	59.6	7.6	1101	12	CNS0039G
8	58.8	7.5	1167	12	CNS07360
c	58.2	7.5	859	12	CNS004YY
9	58	7.4	876	12	CNS009G1
11	57.2	7.3	1092	12	CNS020K7
c	57.2	7.3	1101	12	CNS00EVL
13	57.2	7.3	1101	12	CNS0022U
c	56.6	7.3	1225	12	CNS0161D
15	56.2	7.2	928	12	CNS00DKY
c	56.2	7.2	976	12	CNS04E5M
c	56.2	7.2	1101	12	CNS00E07

18	55.6	7.1	1200	12	CNS016CO
c	55.2	7.1	1101	12	CNS00CYH
20	55	7.1	617	12	BH182965
c	55	7.1	617	12	CNS07080
22	54.6	7.0	820	12	BH1728
23	54.6	7.0	1101	12	CNS00PMC
c	54.2	6.9	475	10	BH14100
25	54.2	6.9	693	9	AV682300
c	54.2	6.9	828	12	CNS011TX
26	53.6	6.9	1084	12	CNS071NH
c	53.4	6.8	507	12	AO181894
28	53.4	6.8	987	12	CNS014FO
c	53.4	6.8	1101	12	CNS003BD
c	53.4	6.8	1190	12	CNS02ON7
c	53.2	6.8	653	12	BH527819
c	53.2	6.8	866	12	AZ668176
34	53	6.8	1101	12	CNS00238
35	53	6.8	1131	12	CNS034FO
36	53	6.8	1201	12	CNS0167M
37	52.4	6.7	914	12	CNS002JY
c	52.4	6.7	1001	12	CNS0155H
c	52.4	6.7	1001	12	CNS0155H
c	52.4	6.7	1061	12	CNS015LM
c	52.4	6.7	1086	12	CNS00YXK
c	52.4	6.7	1101	12	CNS000B8
c	52.2	6.7	859	10	BF274647
c	52.2	6.7	925	12	BH146996
c	52.2	6.7	1101	12	CNS016LI

ALIGNMENTS

RESULT 1

CNS0039G

LOCUS

DEFINITION

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

AL063921

VERSION

AL063921.1

KEYWORDS

GSS.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999)

Genoscope.

BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoss in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp; the same strain used for the BDGP's

PI and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

i. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

FEATURES

source


```

Db 1018 KTGTRTWTKWAAAAA 1034
: : : : : | | |
: : : : : | | |

RESULT 3
CNS01090
LOCUS
DEFINITION
196C24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL167541
VERSION
AL167541.1 GI:7805598
KEYWORDS
GSS: genome survey sequence.
SOURCE
Tetraodon nigroviridis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 524)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 524)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL
Unpublished
REFERENCE
3 (bases 1 to 524)
Genoscope.
DIRECT SUBMISSION
124 a 15 c 24 g 298 t 63 others
/db_xref="taxon:99883"
/clone="196C24"
/clone_lib="G"
/feature="Genoscope sequence ID : COAG196BB12SP1-end :
PUC-ori"
BASE COUNT 124 a 15 c 24 g 298 t 63 others
ORIGIN
1..524
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="196C24"
/clone_lib="G"
/feature="Genoscope sequence ID : COAG196BB12SP1-end :
PUC-ori"
source

Query Match 7.9%; Score 61.4; DB 12; Length 524;
Best Local Similarity 39.1%; Pred. NO. 0.23;
Matches 159; Conservative 43; Mismatches 205; Indels 0; Gaps 0;

QY 39 taaatgtagcacaattttattatgaatagatgaaagctgtagctataaattatgaga 98
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

Db 86 TMDWTWWAAAAASWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 145
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

QY 99 gtaagtattattattatccaaatgtagtctcataatagcataatagcaacttcacta 158
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

Db 146 TTTATTTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTAAATAAATAATATATWTW 205
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

QY 159 aatcttagaataaaaaatgaataatgttaattttttggaggaaatggttaatttttc 218
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

Db 206 TAAAAAATAATAAAAAATTTTTTTTWTWTWTATTTTTTTTTTTTTTTTTTTTTTTT 265
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

QY 219 tacaaaaatgtgtgacagctttacagaccttactcttcacaaattgacttgaacataca 278
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

Db 266 TTTTTTTTTTTTTTTTTTTTTTTTAAATTTTAAATTTTAAATTTTAAATTTAAATAT 325
: : : : : | | | : : : : : | | | : : : : : | | |
: : : : : | | | : : : : : | | | : : : : : | | |

```

DD 1082 AAWAWI'W'I'I'I'MATW'I'WATWMAAAMTTTATATATATATATWAAAWWATTTWAWTATTTW 1023


```

/clone="BACRI9J14"
/note="end : TET3"
BASE COUNT      335 a    54 c    57 g    325 t   105 others
ORIGIN

Query Match          7.4%; Score 58; DB 12; Length 876;
Best Local Similarity 44.6%; Pred. No. 0.8;
Matches 103; Conservative 25; Mismatches 103; Indels 0; Gaps 0;

Qy 1 aattactgtctctaaagtaagcgcttacacccctactaaaaatgtagatacaaaattttat 60
   :||| ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 482 WATWAWTTAWATATATATATWATAAATTTTTAAAAATATATAAAATAAAAHHWWAAAAATAWA 541
   :||| ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 61 tatgaatagatgaagaagcgtgagtacataaaaattatgagagtaaagtttatttatatttc 120
   : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| |
Db 542 AAAATATATANAAAAAATTANAWATAATATTTATATAWAWTWAAAAATTTAAWTAAATTTATATA 601
   : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| |

Qy 121 caaatgtagttcaataagcataatagcaacctcacataaactcttagataataaaaaatgaat 180
   :||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 602 AWAATTTTATATATAANAANAATAAAAAATTAATANAATTTTATATWAWWATATATATATATAT 661
   : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| |

Qy 181 aaatgtttaatttttgaggagaaatggctaatttttctcacaaaattgtgt 231
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 662 WTTANWATAATATTTTAAATATATWTTWAAATATTTNWTAATATATATATATATATTT 712
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
CNS020K7 CNS020K7 1092 bp DNA linear GSS 12-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION
```

sequence:	
ACCESSION	AL175696
VERSION	GI:7813753
KEYWORDS	GSS: genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 1092)
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1092)
AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 1092)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .
FEATURES	Location/Qualifiers
source	1..1092
	/organism="Tetraodon nigroviridis"
	/db_xref="taxon:99883"
	/clone="222L11"
	/clone_lib="G"
	/note="Genoscope sequence ID : C0AG222C06L1-end : T7"
BASE COUNT	383 a 169 c 165 g 262 t 113 others
ORIGIN	

TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

FEATURES Location/Qualifiers
 source 1..1101
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN01007"
 /note="end : SP6"
BASE COUNT 391 a 84 c 76 g 292 t 258 others
ORIGIN

Query Match 7.3%; Score 57.2; DB 12; Length 1101;
 Best Local Similarity 32.1%; Pred. No. 1;
 Matches 164; Conservative 76; Mismatches 269; Indels 2; Gaps 1;
 QY 270 acattacaatcacaaagagggtcctgtttacaaaagaatagtcagaacttcgatattt 329
 Db 520 AAARKTASANNNGASASNNCTTTSSNNACCACCAAAKATANNNNAAAAANNANNTNCCNNGCNN 579
 QY 330 ttgacagtactctttcttaacccttatacccaaatatttaagtgtccatgctctcc 389
 Db 580 NTNNCNTNNNNANKAMAAWAATWTATMMATWTTSAAATSAATSAATTCATTCVTSAA 639
 QY 390 ttatccaactcattgttaactagttttctctgtgagttcctcttgccctataattgaag 449
 Db 640 TTTAAAAAVATTAATT--TTAAATMATTTMTTTTATATATTTTAAAMAATTTTAAAMMTATAT 697
 QY 450 cagttctcgaatacccaactgattttatgaagcccatgctcttttgaaagatttgc 509
 Db 698 AAWTTAAAMAAMWCAATTTTTTTMTATWATAAAACACAACTMTTTTAAAAAACNTS 757
 QY 510 acttcgcttgcactctatttcactgactgactgactgactgactgactgactgactgact 569
 Db 758 AAAAMGSCCTTTTAATTAATVAVCVVBTVCSBSSGNNBNVYVMTTATWATAAAWT 817
 QY 570 tatcagttaggacaatcaaaaagatattagatggcgaggataaatcagaagttact 629
 Db 818 AAWTVAATTNRRYTAMAAVAMHANNNNNWAAWMTTTAAAAAATWAAWTTTATATT 877
 QY 630 gtcaatacaaaagtattgtttatgggtattttatagggtataaatcattactgagcaa 689
 Db 878 TTTDAWMTTATATANTWTTTMTTTTMTTAAWAAWMMWMAAAATTTAAWAAAAATAAA 937
 QY 690 ttccatcatgctttcaattcctcgtggtgaatattggtgactgagactcaaatatt 749
 Db 938 WTWATTWATWATTWTTTAAATTAATWAAATWAAWTTTWTWAAWNTATWATAATWT 997
 QY 750 aaatattggtgtaaaagcgaataaaaaa 780
 Db 998 TWAAWMAAAAAAATATATAAAWMAWMAAAAA 1028

RESULT 14
CNS0161D/c 1225 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106171
VERSION AL106171.1 GI:5620504

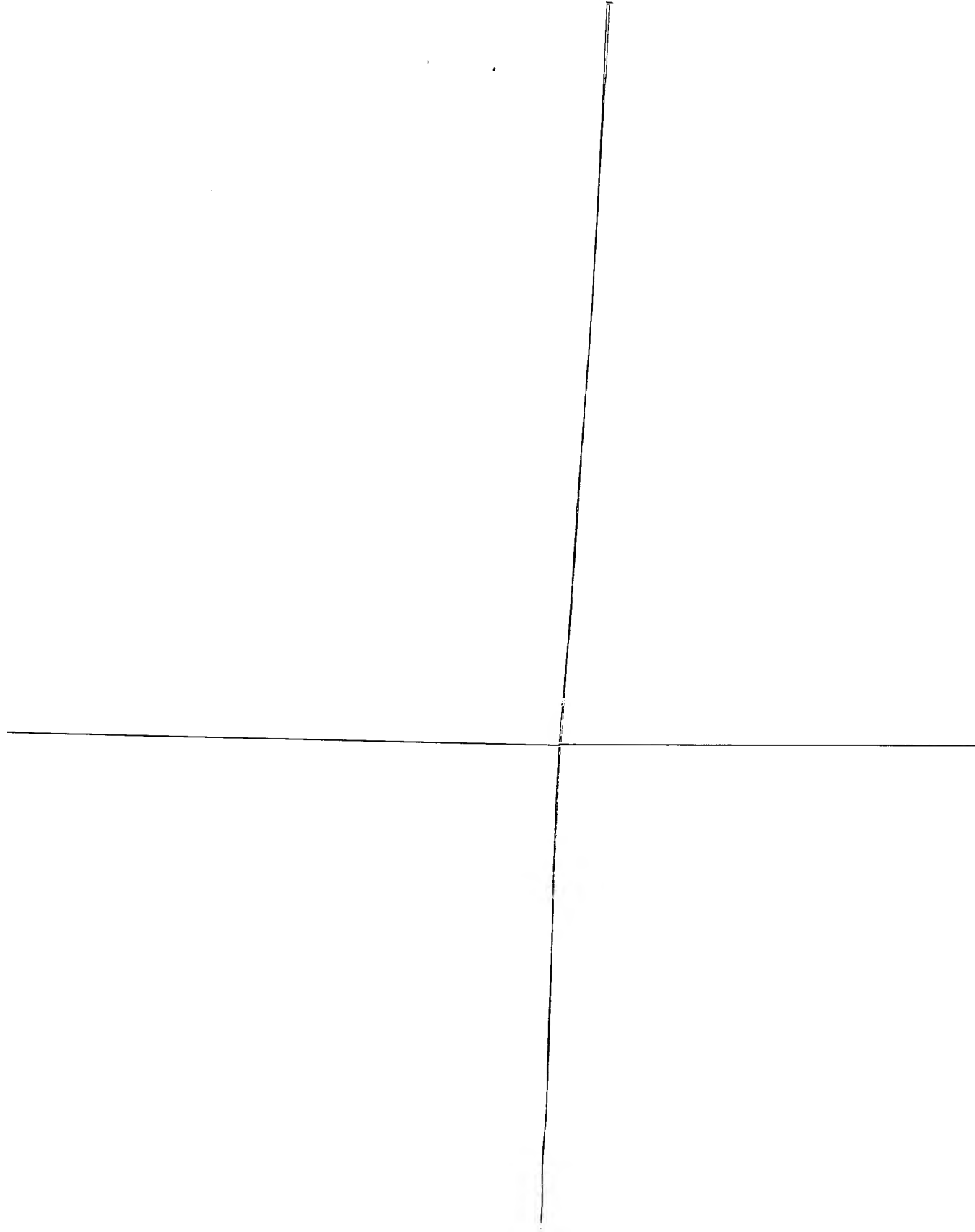
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1225)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

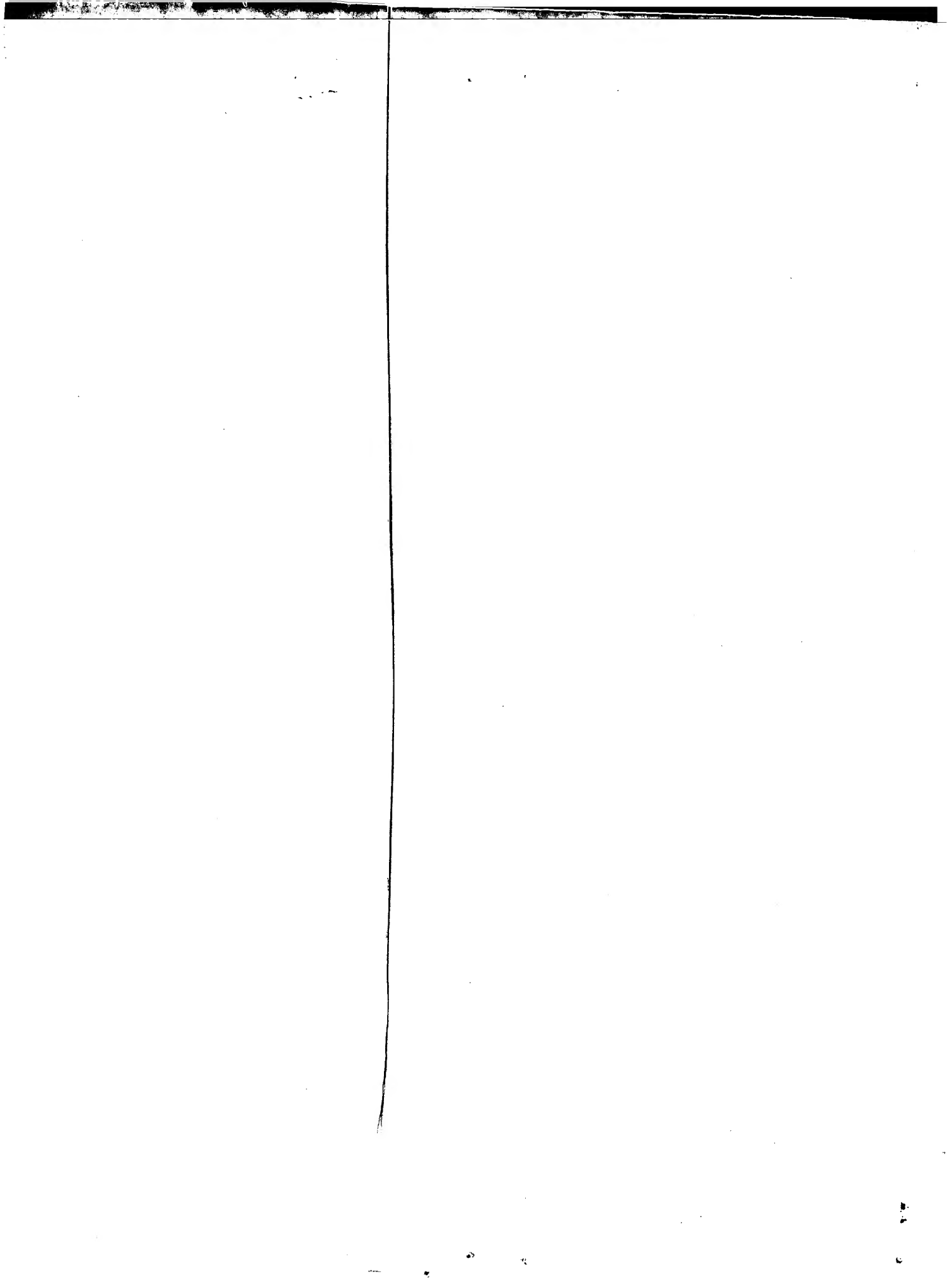
FEATURES Location/Qualifiers
 source 1..1225
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN15C18"
 /note="end : SP6"

BASE COUNT 266 a 128 c 38 g 368 t 425 others
ORIGIN

Query Match 7.3%; Score 56.6; DB 12; Length 1225;
 Best Local Similarity 29.2%; Pred. No. 1.3; Indels 0; Gaps 0;
 Matches 107; Conservative 90; Mismatches 170;
 QY 48 atcaaaattttattatgaatagatgaaagctgtagctataaattatgagtagtaattt 107
 Db 1103 AWWAAWAAWAAWAAWAAWATWTTWTKAATATAAAATATAATATAAAATWAAWWT 1044
 QY 108 ttttattttccaaatgtagtctcataatagcataatagcaacttcactcaaatcttaga 167
 Db 1043 WAWAAAAATATATTTTTTTTTTTTAWAAAAAATAAAATATATWATAAAWWT 984
 QY 168 ataaaaatgaataaattgtaatttttgaggagaaatggttaatttttctcacaaatt 227
 Db 983 AAAAAAATAAATWTTATTTWAAWTTTTTTTAAAAWWTTTTTTTTTTAAWAAWTW 924
 QY 228 gtgtgacagctttacagaccttactcttcacaaatgacttgactgaacattcaacacaaaga 287
 Db 923 WTTATWTTTTRTTRRRARAATWMTTTTTTTTTTTTTTTTTTTTATATTTTTTWHYHY 864
 QY 288 gggctcgtttcaaaaaaatagtcagaacttcataaatttttgcagagctcttttc 347
 Db 863 TYMTTHHHMMMMMMHMMMMGGGGGGTTTTTTTKKTKTKTKTKTKTKTKTKTKTKTK 804
 QY 348 taacctttaccaaataatttaagtgctcatcgtctctcttcttctcaactcattgt 407
 Db 803 ATTTTNTTDTTWAAAAWDDTTWNTKTYMTTHTTKDKMKMKGWNMMWMMGGMTTKT 744
 QY 408 taactag 414
 Db 743 KTTKKDG 737

RESULT 15
CNS00DKY/c 928 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865.1 GI:4948170





TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-883-49

Query Match 2.3%; Score 18; DB 1; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 8

US-08-832-877-49/c
Sequence 49, Application US/08832877
Patent No. 5840506
GENERAL INFORMATION:

APPLICANT: Giordano, Antonio
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 116
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
STREET: Suite 1800 Two Penn Center Plaza
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: US/08/832,877
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-13 US2
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4576 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-832-877-49

Query Match 2.3%; Score 18; DB 2; Length 4576;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 2858 AAGGCAAAAAA 2841

RESULT 9

US-09-128-155-16/c

Sequence 16, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 2.3%; Score 18; DB 3; Length 152331;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 aaggcaaaaaa 780
|||||
DB 8501 AAGGCAAAAAA 8484

RESULT 10

US-08-858-767-21
Sequence 21, Application US/08858767
Patent No. 5837468
GENERAL INFORMATION:
APPLICANT: WANG, Xun
APPLICANT: DUVICK, Jonathan P.
APPLICANT: BRIGGS, Steven P.
TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING
TITLE OF INVENTION: METHOD
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 19-MAY-1997
APPLICATION NUMBER: US/08/858,767
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/481,687
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/325/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136



2
1



```

RESULT 13
US-08-863-028-21
; Sequence 21, Application US/08863028
; Patent No. 5853991
; GENERAL INFORMATION:
; APPLICANT: WANG, Xun
; APPLICANT: DUVICK, Jonathan P.
; APPLICANT: BRIGGS, Steven P.
; TITLE OF INVENTION: PCR-BASED CDNA SUBSTRUCTIVE CLONING
; TITLE OF INVENTION: METHOD
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.

```

